

Koulukorpuksen leksikko:
1.–6.-luokkalaisten aktiivinen sanavarasto

Jaana Laine-Leinonen
Tampereen yliopisto
Kieli-, käännös- ja kirjallisuustieteiden yksikkö
Suomen kieli
Pro gradu -tutkielma
Tammikuu 2013

Tampereen yliopisto
Suomen kieli
Kieli-, käännös- ja kirjallisuustieteiden yksikkö

LAINÉ-LEINONEN, JAANA: Koulukorpuksen leksikko: 1.–6.-luokkalaisten aktiivinen sanavarasto

Pro gradu -tutkielma, 90 sivua + 7 liitettä
Tammikuu 2013

Tutkielmassa tarkastellaan alakoulun 1.–6.-luokkalaisten aktiivista sanavarastoa. Tutkimusaineistona on runsaasta 1 100 Unelmien(i) päivä -aiheisesta kirjoitelmasta koostuva Koulukorpus-aineisto, joka sisältää noin 82 000 sanetta ja vajaat 8 000 lekseemiä. Tutkielma on osa Tampereen yliopistossa suomen kielen oppiaineessa käynnissä olevaa myöhemmän kielenkehityksen tutkimusta. Työn tavoitteena on luoda erityisesti kvantitatiivisin metodein yleiskatsaus Koulukorpuksen leksikkoon ja leksikossa esiintyviin muutoksiin niin eri vuosiluokkien kuin sukupuoltenkin näkökulmasta.

Tutkielma on jaettu kolmeen osaan: Ensimmäisessä osassa tarkastellaan yleisesti saneiden ja sanaluokkien frekvenssejä, eniten käytettyjä lekseemejä sekä niiden frekvenssejä. Toisessa osassa luodaan silmäys yksilövariaatioon vertailemalla muun muassa saneiden sanaluokkien frekvenssejä ja frekvenssien keskihajontaa. Kolmannessa osassa tarkastelun kohteena ovat kerran käytetyt lekseemit eli HL-sanat sekä sanaston rikkaus, jota kartoitetaan erilaisten rikkauslukulaskelmien ja indeksien avulla.

Sanaston ajallista kehitystä havainnoidaan näennäisaikamenetelmällä erityisesti 2., 4. ja 6. vuosiluokkaa vertailemalla. Osa aineiston kirjoitelmista on arvioitu tasonsa perusteella kertomukseksi, skriptiksi, ei skeemalliseksi tai listaksi, ja sanaston monipuolisuutta vertaillaan myös näiden tasoluokitusten näkökulmasta. Tutkimustuloksia verrataan myös muun muassa aikuistenkieleen.

Tutkimustulokset osoittavat, että alakouluikäisten aktiivinen sanavarasto karttuu alakoulun aikana, sillä lekseemien määrä kirjoittajaa kohti kasvaa aina alemmalta vuosiluokalta ylemmälle siirryttäessä. Erityisen paljon kasvaa substantiivilekseemien määrä. Sanaluokkien osuuksissa tapahtuvat muutokset alakoulun aikana ovat enimmäkseen yhden tai kahden prosenttiyksikön luokkaa, eivätkä muutokset tapahdu aina lineaarisesti. Saneiden osalta adjektiivien ja adverbien osuudet kasvavat hieman, kun taas verbien ja pronomien osuudet laskevat. Lekseemien osalta selkeimmin kasvavat verbien, adjektiivien ja adverbien osuudet.

Eniten käytetyt lekseemit ovat kaikilla vuosiluokilla varsin lyhyitä ja konkreettisia, joskin niiden kattavuus koko sanastosta on alemmilla luokilla huomattavasti suurempi kuin ylemmillä. Kumulatiivisen frekvenssin vertailu aikuistenkieleen osoittaa, että Koulukorpus-aineistossa yleisimmät lekseemit kattavat huomattavan suuren osan kaikista aineiston saneista aikuisten kirjoitettuun kieleen verrattuna.

Yksilöllinen variaatio niin aineiden pituuksissa kuin saneiden sanaluokkien osuuksissa on suurta. Saneiden sanaluokkien osuuksien keskihajonta pienenee 2., 4. ja 6. vuosiluokkien tarkastelussa ylemmillä vuosiluokilla, ja hajonta on yleensä aina pienempi tytöillä kuin pojilla. Tyttöjen väliset erot kirjoitelmien sanastossa ovat tästä näkökulmasta pienemmät kuin poikien.

Vain kerran käytettyjä sanoja esiintyy kirjoitelmien lukumäärään suhteutettuna eniten tyttöjen kirjoitelmissa. Niiden osuus saneista on sen sijaan suurempi poikien kirjoitelmissa. Rikkauslukulaskelmien perusteella tytöt käyttävät aina monipuolisempaa sanastoa kuin pojat ja sanasto rikastuu ylemmille vuosiluokille siirryttäessä. Sanasto näyttää myös olevan monipuolisempaa tasoluokitusten kannalta korkeatasoisimmissa aineissa.

Avainsanat: koulu, kirjoittaminen, sanasto, myöhempi kielenkehitys

SISÄLLYS

| | | |
|-------|--|----|
| 1 | JOHDANTO | 1 |
| 1.1 | Tutkimuksen tausta ja aiempi tutkimus | 1 |
| 1.2 | Tutkimusaineisto, -kysymykset ja -hypoteesit | 2 |
| 2 | TUTKIMUKSEN TEOREETTISIA LÄHTÖKOHTIA..... | 5 |
| 2.1 | Sanastontutkimuksesta..... | 5 |
| 2.2 | Alakouluikäisten kieli ja kirjoitustaidon kehittyminen | 7 |
| 3 | KOULUKORPUKSEN LEKSIKKO..... | 11 |
| 3.1 | Yleistä Koulukorpuksen leksikosta | 12 |
| 3.1.1 | Frekvenssit ja lekseemien toisto..... | 13 |
| 3.1.2 | Yleisimmät lekseemit | 17 |
| 3.2 | Sanaluokkien osuudet | 22 |
| 3.2.1 | Substantiivit..... | 27 |
| 3.2.2 | Verbit..... | 33 |
| 3.2.3 | Adjektiivit..... | 42 |
| 3.2.4 | Adverbit..... | 50 |
| 3.2.5 | Pronominit..... | 53 |
| 3.2.6 | Konjunktiot..... | 55 |
| 3.2.7 | Numeraalit | 59 |
| 3.3 | Yksilöllinen vaihtelu ja tasoerot | 62 |
| 3.3.1 | Sisältösanaluokkien ja pronominiin osuudet | 62 |
| 3.3.2 | Substantiivi-verbi-suhde..... | 67 |
| 3.4 | Sanaston rikkaus | 71 |
| 3.4.1 | Yhden kerran esiintyvät sanat | 71 |
| 3.4.2 | Indeksit sanaston monipuolisuuden määrittelyssä | 77 |
| 4 | YHTEENVETO | 86 |

| | |
|---------------|----|
| Lähteet | 91 |
|---------------|----|

Liitteet

1 JOHDANTO

1.1 Tutkimuksen tausta ja aiempi tutkimus

Lasten myöhempää kielenkehitystä ja erityisesti sanastoa on tutkittu Suomessa varsin vähän. Syytä tutkimuksiin olisi, sillä ainakin tutkijoiden keskuudessa ollaan yleisesti huolissaan lasten kirjoitustaitojen kehittymisestä (ks. esim. Pajunen 2012). Erään kyselyn (Opettaja 22–23/2012) mukaan 16 % opettajista pitää lasten ja nuorten kirjoitustaitoa erittäin huonona tai huonona ja jopa 60 % pitää kirjoitustaidon kehittymistä heikentyneenä. Tyypillistä on, että tytöt menestyvät kaikissa äidinkielen taitojen osa-alueiden mittauksissa paremmin kuin pojat (Lappalainen 2008: 48).

Lasten sanavaraston ja sen kehittymisen tutkiminen on erittäin relevanttia. Matilainen (1989: 156) on osoittanut, että lasten sanavaraston laajuus on yhteydessä kirjoitustaitoon. Sanavarastolla on vaikutusta myös lukutaitoon. On nimittäin havaittu, että lapset, joilla on laaja sanavarasto, pärjäävät paremmin luetunymmärtämistehtävissä kuin ne, joiden sanavarasto on suppea, ja toisaalta hyvät lukijat myös saavuttavat laajemman sanavaraston kuin heikommat lukijat (Dockrell & Messer 2004: 45).

Tässä empiirisessä tutkimuksessa pyrin valottamaan alakouluikäisten lasten sanastoa ja siinä luokka-asteittain tapahtuvia muutoksia. Tutkimuksen tavoitteena on luoda yleisesitys Pajusen, Purasen ja Yli-Paavolan Koulukorpus-aineiston (2007–2011) leksikosta. Koulukorpusta on jo aiemminkin käytetty useiden tutkielmien aineistona. Erityisesti Koulukorpuksen leksikkoa käsitteleviä tai sivuavia tutkimuksia ovat:

- Koivunen (2012): Adverbien käyttö alakoululaisten kirjoitelmissa
- Kuisma (2011): Ominaisuuden ilmausten käytön kehittyminen alakouluikäisten kirjoitelmissa
- Mäkelä (2012): Substantiivilausekkeiden kompleksistuminen alakouluikäisten Unelmieni päivä -kirjoitelmissa
- Nahkola (2012): Yhdyslauseiden kehitys kouluikäisen kielessä: tutkimus 1.–6.-luokkalaisten kirjoitelmista
- Palonen (tekeillä): Pronominit alakoululaisten kirjoitelmissa

Tutkimustuloksia esimerkiksi koulukirjoitelmien sanapituuksista ja verbityyppien käytöstä valottaa Pajunen (2012).

1.2 Tutkimusaineisto, -kysymykset ja -hypoteesit

Koulukorpus on koottu 1.–6.-luokkalaisten laatimista Unelmieni päivä -aiheisista kirjoitelmista. Tutkimusaineisto koostuu useaan eri otteeseen kerätyistä kirjoitelmista (ks. aineistolähteet), joita nimitetään tässä tutkimuksessa yhteisnimellä Koulukorpus 2007–2011. Pajusen, Purasen ja Yli-Paavolan 1.–6.-luokkalaisten kirjoitelmista koostuva alkuperäinen Koulukorpus 2007 kattaa tutkimusaineistosta 71 000 sanetta. Sitä on täydennetty Raahen aineistolla, joka koostuu 1.–6.-luokkalaisten kirjoitelmista, 3. ja 4. luokan kirjoitelmia sisältävällä kielikylpyaineistolla, kielikoulun 3.-luokkalaisten aineilla ja Sintosen keräämillä 5.-luokkalaisten kirjoitelmilla. 2.-luokkalaisten aineistoa on täydennetty myös kirjoitelmilla, jotka on kerännyt Honko (2007 ja 2008). Lisäaineistot nostavat Koulukorpuksen sanemäärän yli 82 000:een. Kaikki aineistoon kuuluvat aineet on kirjoitettu samasta aiheesta, jonka otsikkona on ollut Unelmieni päivä tai Unelmien päivä (Hongon aineisto).

Suurimmalle osalle tutkimusaineistosta on tehty peruskoodaus (mm. oppilaat, koulut, alkuperäinen teksti ja korjattu teksti, lekseemit ja sanaluokkakategoriat), mutta olen jonkin verran täydentänyt ja tarkistanut koodauksia. Osa kirjoitelmista, Pajusen, Purasen ja Yli-Paavolan keräämät, on myös luokiteltu siten, että niiden tasoksi on arvioitu joko lista, ei skeemaa, skripti tai kertomus (ks. esimerkit Pajunen, 2012). Tasoluokituksesta on huomioitava ainakin se, että esimerkiksi 4.-luokkalaisten listat saattavat poiketa suuresti 2.-luokkalaisten listoista ja että ei skeemaa -tyyppiset kirjoitelmat ovat rakenteeltaan hyvin heterogeenisiä. Lisäksi tasoluokitukset vaativat vielä tarkistamista, joten joidenkin kirjoitelmien tasoluokitus saattaa muuttua, kun tasoluokitusten kriteerit tarkentuvat. Tutkin kirjoitelmien leksikkaa hieman myös tasoluokitusten näkökulmasta, jolloin kaikki tasoltaan luokittelematon aineisto jää tarkastelun ulkopuolelle. Aineistoa rajataan muutenkin aina tarpeen mukaan.

Kielikylpyaineistoa sisältyy 3. ja 4. luokan aineistoihin. Sen osuus 3. luokan saneista on 18,3 % ja 4. luokan saneista 23,0 %. Kielikouluaineiston osuus 3. luokan saneista on 22,8 %. Kielikouluaineistossa merkillepantavaa on kirjoitelmien hyvä taso, mikä saattaa aiheuttaa vinoumaa 3. luokan tuloksiin. Toisaalta 3. luokan kielikylpyaineisto saattaa olla keskimääräistä heikotasoisempaa, mikä osaltaan lisää 3. luokan aineiston heterogeenisyyttä.

Tutkin sanastoa välillä koko aineistosta, välillä yksittäisten luokkien (tavallisimmin 2., 4. ja 6.) tai erikseen tyttöjen ja poikien kirjoitelmista. Toisaalta tutkimukseni tarkoituksena on siis valottaa koko Koulukorpuksen leksikkoa, toisaalta osoittaa myös mahdollisia sanastossa tapahtuvia muutoksia oppilaiden varttuessa sekä tyttöjen ja poikien välillä. 2., 4. ja 6. luokan aineistojen vertailu riittänee kuvaamaan varttumisen myötä tapahtuvaa kehitystä. Lisäksi vertaan Koulukorpuksen leksikkoa joiltakin osin opetus- ja kulttuuriministeriön hallinnoiman Tieteen tietotekniikan keskus CSC:n internet-sivuilta löytyvään Suomen sanomalehtikielen taajuussanastoon (2004), lasten kuvakirjojen (Peltola 2011) ja peruskoulun 2. ja 3. luokan oppikirjojen sanastoon (Jaakola 2004) sekä Opettajan sanastokirjan (1970) esittämiin tuloksiin.

Analyysia varten käytän enimmäkseen koko aineistoa, mutta joidenkin muuttujien tarkastelussa jätän esimerkiksi numerot tai tasoluokittelemattoman analyysin ulkopuolelle. Näistä toimenpiteistä mainitsen aina erikseen. Alun tarkistusten sekä mahdollisten rajaustoimenpiteiden jälkeen käsittelen tutkimusaineistoani kvantitatiivisesti. Tulokset esitän usein taulukoiden ja kuvioden avulla. Tavoitteenani on jäsentää leksikkoa siten, että sitä olisi helppo käyttää vertailuaineistona lapsenkielentutkimuksessa. Tavoitteenani on myös kartoittaa sitä, kuinka leksikko muuttuu luokka-asteelta toiselle siirryttäessä ja kuinka kirjoitelman taso, pituus ja lekseemien määrä korreloivat keskenään ja kuinka monipuolista kirjoitelmien sanasto on.

Yleisen kuvauksen ohella pyrin vastaamaan ainakin seuraaviin kysymyksiin:

1. Mitä lekseemejä koululaisten kirjoitelmissa esiintyy eniten? Käyttävätkö tytöt ja pojat erilaisia lekseemejä?
2. Mitkä ovat eri sanaluokkien osuudet kouluikäisten kirjoitelmissa?
3. Kuinka leksikko eroaa eri luokka-asteiden välillä? Entä tyttöjen ja poikien välillä?
4. Minkälaista alakouluikäisten sanasto on verrattuna lasten kuvakirjojen, 2.-3.-luokkalaisten oppikirjojen tai sanomalehtikielen sanastoon? Entä 40 vuoden takaisiin tutkimustuloksiin kouluikäisten kielestä (Opettajan sanastokirja 1970)?
5. Minkälaista yksilöiden välistä variaatiota sanastossa muun muassa voidaan havaita?
6. Korreloivatko kirjoitelmien taso ja sanaston monipuolisuus keskenään?
7. Mitä rikkauslukuindeksit kertovat kouluikäisten sanaston monipuolisuudesta?

Tutkimushypoteesini ovat, että sanasto monipuolistuu lasten varttuessa ja että tyttöjen käyttämä sanasto on samanikäisten poikien sanastoa monipuolisempaa.

Tuloksiani esitän sanalistojen muodossa sekä numeerisesti taulukoiden ja diagrammien avulla. Sana- ja lause-esimerkkien avulla pyrin selittämään tuloksia sekä myös tuomaan esiin kiinnostavia ilmiöitä. Jotta tutkimus pysyisi laajuudeltaan gradukontekstiin sopivana, esimerkiksi semanttista analyysia ei juurikaan tehdä. Tutkimuskohteena on lisäksi usein vain kolme luokka-astetta, 2., 4. ja 6. luokka, sillä kyseisten luokka-asteiden tarkastelu antaa riittävän hyvän kuvan alakoulun aikana sanastonkäytössä tapahtuvista muutoksista ja kehityksestä.

Kuten yleensä, myös Koulukorpuksen kirjoitelmien sanastoon vaikuttavat yleisesti kirjoitelman aihe, oletettu lukija – ehkäpä opettaja – ja käyttötarkoitus. Kirjoitelmien sanasto ei kerro lapsen koko sanastosta, sillä lapsilla on sekä aktiivinen, itsensä käyttämä, että passiivinen sanavarasto, joka kehittyy hieman aktiivista nopeammin (Saarinen 1982: 62). Koulukirjoitelmien kautta pääsee käsiksi vain osaan koululaisten aktiivisesta sanavarastosta. Sen vuoksi tämänkaltaisen tutkimuksessa on oleellisista osoittaa sanastossa tapahtuneita muutoksia eikä vain luetella käytettyjä sanoja. Myös sanaston vertailu muihin lähteisiin hahmottaa lisää kouluikäisten sanastoa, mutta todella vertailukelpoista aineistoa on vaikea löytää siksi, että kirjoitelmien sanasto on riippuvainen aihepiiristä ja tekstilajista.

2 TUTKIMUKSEN TEOREETTISIA LÄHTÖKOHTIA

2.1 Sanastontutkimuksesta

Tässä tutkimuksessa keskitytään alakouluikäisten leksikkoon, joka koostuu lekseemeistä. Lekseemi on nimitys sanalle sanaston yksikkönä; lekseemi edustaa kaikkia yhden sanan taivutusmuotoja. Tavallisessa puheessa tai kirjoitetussa tekstissä esiintyvät sanat ovat sanamuotoesiintymiä eli saneita. Ne ovat lekseemin muunnoksia, jotka ovat olemassa syntaksin ja morfologian ominaisuuksien takia. Lekseemit ovat muodostavat kielen sanaston ytimen ja niitä käytetään sanakirjan sanoina. Sanakirjassa lekseemien sananmuotoja edustaa sanakirjamuoto, joka on suomessa nominien kohdalla yksikön nominatiivi ja verbien kohdalla A-infinitiivi. Niiden tärkein ominaisuus on niiden sanaluokkakategoria. Lekseemiin kuuluu ainakin yksi denotaatio, homonyymien kohdalla useampia. (Karlsson, 2006: 83–86.) Sane ja lekseemi ovat tämän tutkimuksen kannalta keskeiset yksiköt.

Niemikorpi (1991: 14) on luetellut useita sanaston yksikön määrittelyyn vaikuttavia kysymyksiä: Niihin kuuluu esimerkiksi sanan käsitteen ongelma eri kielissä, leksikon perusyksikön ongelma, kuten mikä yleensä on sana sekä perusyksikön suhde johdoksiin, yhdyssanoihin ja kiinteisiin yhdistyksiin sekä lyhenteisiin, sekä tekstissä olevat matemaattisten ja graafisten symbolien käsittely. Jaakola (2004: 12) jättää huomiotta numerot, kirjainsymbolit, lyhenteet ja epätäydelliset sanat.

Tämän tutkimuksen aineistossa omiksi lekseemeikseen on määritelty muun muassa yhdyssanat, johdokset, lyhenteet ja selvästi leksikaalistuneet tai kiteytyneet adverbit, kuten esimerkiksi *sti*-päätteiset adjektiivijohdokset sekä tapaa tai paikkaa ilmaisevat lekseemit *hyvin*, *ostoksilla* ja *kotona*. Mikäli lekseemeiksi arvioitaisiin vain sanojen ”juuret”, antaisi tutkimus kouluikäisten kielestä huomattavasti todellisuutta köyhemmän kuvan.

Numeroiden käsittely aineistossa ei ole aivan yksioikoista. Esimerkiksi Jaakola (2004) jättää numerot pois, sillä hänen tutkimuksensa aineisto koostuu oppikirjoista, joista esimerkiksi luonnontieteellisissä ja matematiikan oppikirjoissa numeroja esiintyy valtavasti. Numeroiden jättäminen aineistoon aiheuttaisi sen, että lekseemien määrä näyttäisi huomattavasti suuremmalta ja leksikko rikkaammalta kuin on, sillä onhan numeroita ainakin teoriassa loputon määrä. Tässä tutkimuksessa numerot ovat kuitenkin oleellinen osa kirjoitelmien tekstiä – nehan voisi aina kirjoittaa myös kirjaimin. Tässä tutkimuksessa nojataan myös varsin pitkälti Koulukorpuksen peruskoodaukseen koodauksen ongelmia

ratkaistaessa, ja numerot on koodattu omiksi numeraalien kategoriaan kuuluviksi lekseemeikseen. Numerot jätetään analyysin ulkopuolelle kuitenkin silloin, kun ne vaikuttaisivat liikaa sanaston monimuotoisuuteen ja antaisivat turhan ruusuisen kuvan sanaston rikkaudesta esimerkiksi yhden kerran käytettyjen sanojen eli HL-sanojen tarkastelussa.

Mikä on sanastontutkimuksen kannalta merkittävää? Sanastontutkimuksessa merkittäviä kvantitatiivisia ominaisuuksia ovat esimerkiksi kielen ekonomia ja rikkaus, luettavuus, tiiviys ja kumulatiiviset ominaisuudet kuten ilmausten keskipituudet (Niemikorpi 1991: 24). Koulukorpus-aineiston tekstisanojen sananpituuksia, sanemääriä ja kirjoitelmien tasoja sekä erilaisten sanarakenteiden tekstisanapituuksia on jo tutkittu (ks. Pajunen 2012: 8–18). Tässä tutkimuksessa keskitytään leksikkoon esimerkiksi sana-sane-suhteen, sanaluokkien osuuksien, erilaisten lekseemien sekä kielen rikkauden osalta.

Kouluikäisten sanastosta on Suomessa tehty aiemmin vain vähän tutkimusta. Opettajan sanastokirja (Karvonen & al. 1970) on ensimmäinen suomalainen sanastotutkimus lasten aktiivisesta sanavarastosta. Sen aineisto, oppilaiden kirjoittamat aineet, on kerätty yli 8700 3. ja 4. luokan oppilaalta silloisen Keski-Suomen läänin kansakouluista. Oppilaat kirjoittivat aineensa monenlaisista aiheista. Aiheet arvottiin koulujen kesken. Tutkimuksen tuloksena esitellään sanalistoja sekä 600 yleisimmistä sanasta että kaikista aineissa käytetyistä sanoista. Murteelliset ja omintakeiset sanat esitetään vielä omissa listoissaan. Sanaluokkien osuuksia ja tiettyjen lekseemien käyttöä verrataan tutkimuksessa sekä sosiaaliluokkien että 3. ja 4. luokan kesken. Näitä tuloksia hyödynnetään myös tässä tutkimuksessa arvioitaessa, millä tavalla sanaluokkien osuudet voisivat kuvastaa kirjoittajan kehitystä. Sanalistoja ei kuitenkaan verrata, sillä ne eroavat toisistaan paljon jo kirjoitelmien aiheiden erilaisuuden vuoksi.

Saarela (1997) on tutkinut peruskoululaisten kirjoitelmien kehittymistä sanastotutkimuksen valossa. Hän on kerännyt tutkimusmateriaalinsa 1980-luvulla yhteensä 69 oppilaalta. Kyseessä on pitkittäistutkimus, ja Saarela on lopulta ottanut analyysiinsä mukaan vain ne oppilaat, jotka ovat olleet läsnä jokaisessa kirjoitustilanteessa. Oppilaat ovat kirjoittaneet 2. ja 6. luokalla aineen aiheesta Aivan tavallinen päivä ja 4. ja 8. luokalla aiheesta Näin aion viettää kesäni. Tutkimusaineistostaan Saarela on analysoinut muun muassa aineiden pituuden ja sanaston kehittymistä, sanaluokkien osuuksia sekä sanaston abstraktistumista ja asennetason muutoksia, ja hän tekee huomioita yksilön kehityksestä sekä vertailee tuloksia eri

luokka-asteiden että tyttöjen ja poikien välillä. Analyysinsä välineenä hän käyttää muun muassa Shannon-Weaverin diversiteetti-indeksiä eli Shannonin entropiaa, mihin palataan tässä tutkimuksessa luvussa 3.4.2 Indeksit sanaston monipuolisuuden määrittelyssä.

Tässä tutkimuksessa keskitytään kouluikäisten käyttämiin lekseemeihin sananmuotoesiintymien sijasta. Käytössä oleva menetelmä on näennäisaikamenetelmä, jossa leksikkoa tarkastellaan sekä kokonaisuutena että eri luokka-asteiden näkökulmasta. Aineisto on enimmäkseen valmiiksi koodattua. Koodaus sisältää muun muassa lekseemien alkuperäisen ja korjatun kirjoitusasun, lekseemi, sanaluokan, yhdyssanavirheet, tiedot koulusta, oppilaan sukupuolesta, sekä joidenkin kirjoitelmien osalta kirjoitelman tason. Oppilaat on identifioitu siten, että esimerkiksi 4.-luokkalaisten pojan yksilöity koodi voi olla ”kaupunki 4P144”. Poikaa merkitsee P-kirjain, tyttöä vastaavasti taas T.

2.2 Alakouluikäisten kieli ja kirjoitustaidon kehittyminen

On yleinen käsitys (ks. esim. Leiwo 1980: 103), että kielen osaamiselle ei ole mitään ylärajaa. Yleisesti hyvä kielitaito voidaan määritellä kyvyksi sopeutua erilaisten tilanteiden kielenkäytölle asettamiin vaatimuksiin. Kielitaidon koko elämänmittainen kehitys näkyy tavallisesti ainakin sanaston karttumisena.

Lapsi ajattelee egosentrisesti 7–8 vuoden ikäiseksi. Sen jälkeen alkaa pikku hiljaa siirtyminen sosiaalistuneeseen ajatustapaan, johon kuuluu vähemmän itsekeskeinen ajattelu sekä kuvakaavojen ja analogisten kaavojen käyttö ajattelun rakentumisessa. (Piaget 1973:123.) Aikakäsite sekä nopeuskäsite, eli tietyn matkan kulkeminen tietyssä ajassa, kehittyy vasta, kun lapsi on noin kahdeksanvuotias. Ajanilmausten käytön pitäisi siis lisääntyä ensimmäisen kouluvuoden jälkeen.

Lapsen ajattelu on konkreettista noin 7–8-vuotiaasta 11–12 vuoden ikäiseksi, minkä jälkeen alkaa abstraktisten operaatioiden vaihe. Tämä niin sanottu konkreettisten operaatioiden kausi voidaan jakaa kahteen vaiheeseen, joista varhaisempi on yksinkertaisten operaatioiden vaihe ja toinen kokonaisjärjestelmien valmistumisvaihe, johon kuuluvat tilan ja ajan käsitteiden hahmottaminen. (Piaget 1988: 106.) Lapsen ensimmäiset kouluvuodet kuuluvat siis konkreettisten operaatioiden kehityskauteen. Vasta kymmenennen ikävuoden jälkeen lapset osaavat yleensä muotoilla abstrakteja selityksiä abstrakteille, irrallisille sanoille (johdonmukainen, mukiinmenevä), runosäkeille tai sananparsille, mutta konkreettisia

selityksiä voi esiintyä vielä viisitoistavuotiaillakin (Saarinen 1982: 62). Kun abstraktisten operaatioiden vaihe on saavutettu, lapsi pystyy esimerkiksi oppimaan uusia käsitteitä ilman konkreettisia esimerkkejä. Uusien sanojen oppiminen vanhoihin sanoihin suhteuttamalla helpottuu. Konkreettisten operaatioiden vaiheessa uusien käsitteiden oppiminen vaatii vielä niiden yhdistämistä lapsen konkreettiseen kokemusvarastoon. (Karvonen & al. 1970: 14.)

Berman (2004: 12–13) jakaa kielenomaksumisen kolmeen vaiheeseen, esikieliopilliseen, kieliopillistuneeseen ja konventionaalistuneeseen vaiheeseen. Näistä esikieliopillinen vastaa Piaget’n konkreettisten operaatioiden kehityskautta. Kieliopillistuneessa vaiheessa keskeisessä asemassa ovat kielen rakenteet ja säännöt ja konventionaalistuneessa vaiheessa taas sidonnaisuus kontekstiin ja diskurssiin. Kouluikäisten kirjoitelmissa kehitysvaiheiden pitäisi näkyä niin, että abstrakteja ilmauksia, vertauskuvia tai metaforia alkaa esiintyä vasta neljäsluokkalaisten ja sitä vanhempien kirjoitelmissa, sillä abstraktimpien ilmausten käyttö vaatii ymmärtämystä siitä, että sanalla voi olla useampi kuin vain yksi merkitys (Tolchinsky 2004: 237). Nuorimmilla kirjoittajilla sanasto on siis enimmäkseen varsin kouriin tuntuva, vaikka yksilöllistä vaihtelua onkin.

Kouluiässä lasten sanasto kasvaa ja sanojen merkitykset monipuolistuvat lisääntyvien kokemusten myötä. Varttuessa eteen tulevat uudet tilanteet asettavat uusia kielellisiä vaatimuksia. Koulussa lapset oppivat kirjallisen viestinnän, lukemisen ja kirjoittamisen perustaidot. (Leiwo 1980: 104.) Sanasto rikastuu koulussa kaikkien oppiaineiden tunneilla, samoin koulujen ulkopuolisissa yhteisöissä ja yhteyksissä, joissa oppija joutuu kieltä käyttämään. Sanaston karttuminen on sekä kvantitatiivinen että kvalitatiivinen ilmiö. Sen lisäksi, että sanoja tulee lisää, ne myös tarkentuvat ja siirtyään semanttisesti perustasolta alemmille tasoille: puusta tulee mänty, koivu tai kuusi, kukasta ruusu, kielo tai sinivuokko. (Koppinen 1982: 92–93.) Iän karttuessa yksilöiden kielen välillä havaittavat erot myös kasvavat, sillä kieli on herkkä sosiaalisille ja kulttuurisille vaikutteille (Tolchinsky 2004: 235). Persoonallinen itseilmaisun tapa saa muodon kypsyymisen myötä ja siihen vaikuttaa niin sosiokulttuurinen tausta kuin elämäkokemuskin (Berman 2004: 15). Lapsen passiivisella sanavarastolla on havaittu olevan merkittävä vaikutus kirjoittamistaitojen kehitykselle ainakin oikeinkirjoituksen, kirjoitelman rakenteen ja sisällöllisen rikkauden sekä käytettyjen hahmottamisstrategioiden osalta (Matilainen 1989: 149–157).

Koulutuksella on tärkeä rooli lapsen myöhemmässä kielenkehityksessä, jonka kaudelle sijoittuu myös luku- ja kirjoitustaidon saavuttaminen (Nippold 2004: 1–5). Sanavaraston niin laadullista kuin määrällistäkin kehittämistä voi pitää yhtenä opetuksen

tärkeimmistä tehtävistä (Matilainen 1989: 168). Kouluvuosien aikana lapset oppivat paljon uutta sanastoa niin koulutuksen kuin vapaa-ajankin tarpeisiin. Uuden sanan oppiminen vaatii sitä, että lapsi oppii yhdistämään luonnollisen asiantilan, muodon ja merkityksen. Uusi sana on myös liitettävä jo aiemmin opittuun leksikkoon. Se, kuinka helposti lapsi uusia sanoja omaksuu, riippuu ensimmäisinä kouluvuosina kontekstista sekä jo valmiiksi omaksutusta sanastosta. Lapsen varttuessa sanavaraston jatkuva monipuolistuminen edesauttaa uusien sanojen omaksumista, eikä kontekstin merkitys sanan omaksumisessa ole enää niin suuri. (Dockrell & Messer 2004: 36–37, 43.)

Kuten puhutun kielen kehitys, myös kirjoitetun kielen ja lausetajun kehitys noudattavat selviä säännönmukaisuuksia. Esi- ja alakoululaisten puhutulle kielelle tyypillinen ilmiö on niin sanottu ketjuvirke, jossa irralliset lauseet yhdistetään toisiinsa löysäksi kokonaisuudeksi käyttämällä erilaisia konnektoreja, esimerkiksi *ja*, *sitten* ja *tai*. Ensimmäisinä kouluvuosina tällaiset ketjuvirkkeet ovat yleisiä myös lapsen kirjoitelmissa, mutta katoavat yleensä viimeistään kolmannen kouluvuoden aikana. Samalla lapsi oppii sulauttamaan enemmän informaatiota yhteen kielelliseen yksikköön. (Leiwo 1980: 109–110.) Vanhemmat lapset pystyvät luomaan tiiviimpiä ja paremmin jäsenneltyjä tekstejä ja heidän taitonsa yhdistellä lauseita erilaisilla rakenteilla kasvaa (Tolchinsky 2004: 237).

Kieliopillinen kehitys on esimerkiksi taivutuksen osalta loppuunsaatettu koulun alkuun mennessä, vaikka monimutkaisen rakenteet tuottavat vaikeuksia vielä alaluokilla. Koulu asettaa lapselle aivan uusia kielenkäytön valmiuksia, kuten erilaiset rekisterit ja muut julkisen kielen normit. Uusiksi kielenkäytön alueiksi muodostuvat lukeminen ja kirjoittaminen. Kirjoittaminen edellyttää lapselta tiettyjä kognitiivisia kykyjä, tekstistrategioitten ja kielen normien hallintaa. (Leiwo 1980: 109.)

POPS2004:n asettamat tavoitteet vuosiluokkien 1 ja 2 kirjoitustaidolle ovat ensinnäkin kirjoittamisen perustekniikoiden hallinta, mihin kuuluu keskeisenä sisältönä erityisesti 1.-luokalla oikea kynäote sekä kirjaimet. Näiden jälkeen tavoitteena on oikeinkirjoituksen hallinta sana- ja lausetasolla, mikä pitää sisällään ainakin sanavälit ja sanan jakamisen eri riveille, isot alkukirjaimet tutuissa nimissä ja lauseiden alussa, lauseiden lopetusmerkit sekä niiden käytön omissa teksteissä. Tekstin tuottamisen ilo on tärkeintä. Tuotettavat tekstit perustuvat omiin arkikokemuksiin, havaintoihin, mielipiteisiin ja mielikuvitukseen. 3.–5. vuosiluokilla tavoitteena on, että oppilaan taito tuottaa erilaisia tekstejä kehittyä. Keskeisiin sisältöihin kuuluu edelleen tutusta asiasta kirjoittaminen, juonellinen kertomus sekä

mielipiteen ilmaisu ja perusteleminen. Vuosiluokkien 6–9 tavoitteena on, että oppilas tottuu suunnittelemaan viestintäänsä ja pyrkii etenemään kirjoitustehtävissään tavoitteellisesti. Oppilas osaa myös ottaa huomioon vastaanottajan ja kirjoitustehtävän asettamat vaatimukset tekstiä tehdessään. Keskeisiin sisältöihin kuuluu edelleen erilaisten tekstien tekeminen, suunnittelu sekä lause- ja virketason oikeinkirjoituksen vakiinnuttaminen sekä tekstin jäsentelyn opettelu. (POPS2004: 23–30.) Sintonen (2011) on tutkinut pro gradu -työssään suunnittelemisen vaikutusta 5. ja 7. -luokkalaisten kirjoitelmiin. Hänen mukaansa suunnitelman teettäminen parantaa tekstin koherenssia ja sisällön runsautta, pienentää eroa sukupuolten ja yksilöiden välillä, mutta toisaalta vähentää persoonallista otetta ja tunteiden ilmaisemista. Sintonen otaksuukin, että suunnittelu vie ainakin 5.-luokkalaisilta vielä niin paljon huomiota, että lopullisen tekstin tekemiseen ei riitä enää voimia yhtä paljon kuin ilman suunnitelmaa. Tämän tutkimuksen aineistoon sisältyy myös Sintosen kokoamaa ja käyttämää aineistoa, mutta suurin osa aineistosta koostuu kirjoitelmista, joita ei ole ainakaan ohjatusti suunniteltu.

3 KOULUKORPUKSEN LEKSIKKO

Tässä luvussa alakouluikäisten sanastoa pyritään esittelemään useista näkökulmista. Alaluvussa 3.1 luodaan yleisluontoisia katsauksia koko leksikkoon ja frekventeimpiin lekseemeihin. Aluksi tutkimuksen kohteena ovat frekventeimpien lekseemien sanaluokat, niiden kumulatiivinen frekvenssi sekä sanojen keskimääräinen toisto. Koko aineiston yleisimmät lekseemit (kuvio 5 ja liite 2) sekä niiden frekvenssi luokittain tuodaan näkyville kuviossa. Sanaston kehitystä seurataan enimmäkseen vertailemalla keskenään 2., 4. ja 6. luokan aineistoja. Vertailua muihin aineistoihin tehdään lähes jokaisen tutkittavan joukon yhteydessä.

Alaluvussa 3.2 tarkastellaan ensin yleisesti kirjoitelmien saneiden ja lekseemien sanaluokkien osuuksia, sen jälkeen tutkimuksen kohteena ovat erikseen yleisimmät substantiivit, verbit, adjektiivit, pronominit, adverbit, numeraalit ja konjunktiot. Näkökulma vaihtelee hieman sanaluokasta riippuen. Yleisimmät substantiivi-, verbi-, adjektiivi- ja adverbilekseemit on esitetty sekä kuvioden avulla että liitteissä. Eniten keskitytään substantiivi-, verbi- ja adjektiivilekseemeihin, joiden käyttöä tarkastellaan koko aineiston ohella erikseen tyttöjen ja poikien kirjoitelmissa eri vuosiluokilla. Pronomini- ja adverbilekseemeistä sekä konjunktioista esitetään yleisimmät lekseemit sekä niiden osuudet. Lisäksi niiden käyttöä on arvioitu Opettajan sanastokirjan esittämien argumenttien valossa. Konjunktioiden käyttöä on myös verrattu CSC:n Suomen sanomalehtikielen taajuussanastoon. Numeraaleista on nostettu esiin muutamia aineistosta esiin tulevia näkökohtia.

Eri sanaluokista on myös tehty tai tekeillä syväluotaavampia tutkimuksia, joten esimerkiksi semanttisiin näkökantoihin tartutaan tässä tutkimuksessa hyvin vähän.

Alaluvussa 3.3 pyritään tuomaan esiin sanaston vaihtelua yksilötasolla substantiivi-verbi-suhdetta sekä sanaluokkien osuuksia vertailemalla. Olennaisinta tässä luvussa on se, kuinka vaihtelevia kirjoitustyyli ja -taito ovat yhden ikäluokan sisällä.

Analyysiosan lopussa alaluvussa 3.4 perehdytään yhden kerran esiintyviin eli hapaks legomenom -sanoihin, lyhyemmin HL-sanoihin, sekä käytetään valikoituja indeksejä ja kaavoja kirjoitelmien sanaston rikkauden määrittämiseen. Sanaston rikkautta arvioidaan sekä vuosiluokittain, tasoluokittain että tyttöjen ja poikien kirjoitelmista erikseen.

3.1 Yleistä Koulukorpuksen leksikosta

Taulukossa 1 on esitetty perustietoja tässä tutkimuksessa käytetystä aineistosta. Taulukon tiedot on koottu koko aineistosta, joka sisältää Pajusen, Purasen ja Yli-Paavolan Koulukorpuksen lisäksi myös muut käyttämäni aineistot. Lisätietoja, muun muassa saneiden ja lekseemien määrät, löytyy liitteenä 1 olevasta taulukosta.

Taulukko 1. Aineiston tunnuslukuja.

| | 1. luokka | 2. luokka | 3. luokka | 4. luokka | 5. luokka | 6. luokka | koko aineisto |
|---------------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|---------------|
| oppilaita | 123 | 224 | 207 | 215 | 182 | 160 | 1111 |
| tyttöjä | 63 | 115 | 99 | 118 | 90 | 78 | 563 |
| poikia | 60 | 109 | 108 | 97 | 92 | 82 | 548 |
| saneita yhteensä noin | 2697 | 10466 | 13310 | 18829 | 19389 | 17578 | 82269 |
| lekseemejä yhteensä noin | 724 | 1882 | 2431 | 3015 | 3178 | 3007 | 7770 |
| aineiden pituus keskimäärin (sanetta) | 22 | 47 | 66 | 88 | 107 | 110 | |
| aineiden pituuden keskihajonta | 18 | 26 | 39 | 46 | 56 | 59 | |
| psin aine (sanetta) | 94 | 134 | 236 | 233 | 329 | 346 | |
| lyhin aine (sanetta) | 1 | 4 | 5 | 8 | 17 | 6 | |

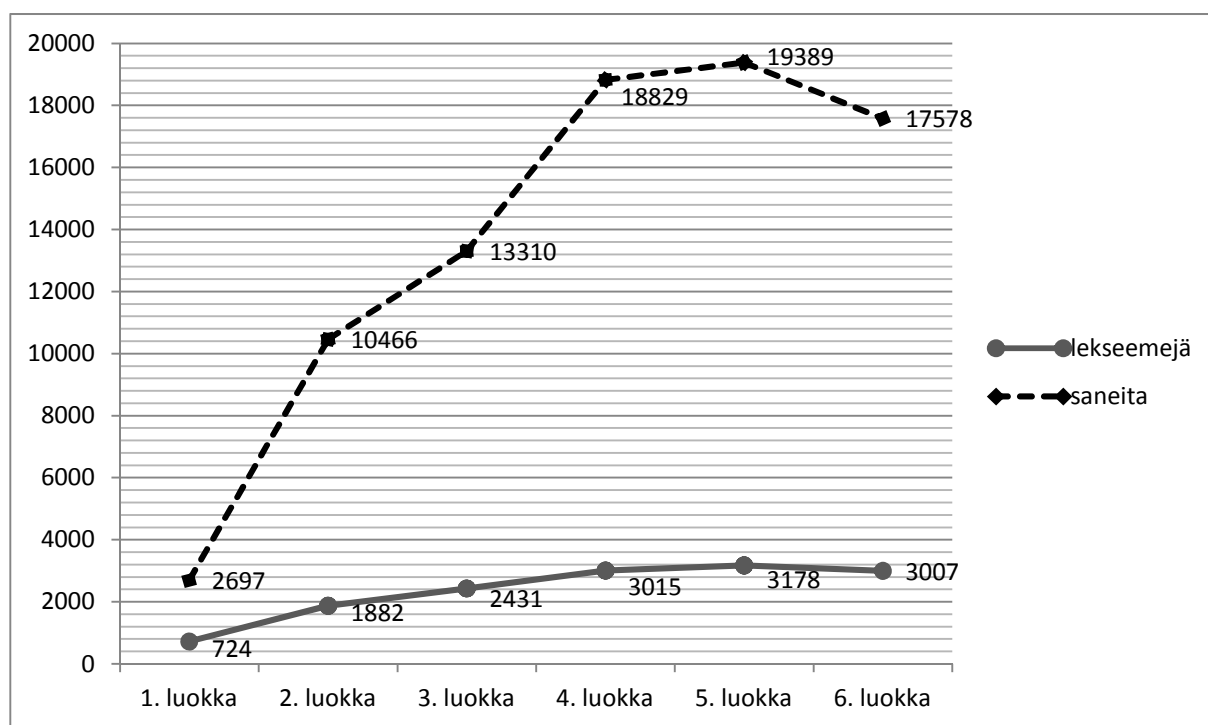
Aineiden pituuden keskihajonnan perusteella voi päätellä, että erot kirjoitelmien pituudessa kasvavat luokka-asteelta ylemmälle siirryttäessä. Myöhemmässä analyysissä hahmottuu hieman sekin, kasvavatko yksilöiden väliset erot muutenkin ylemmillä vuosiluokilla.

Analyysissä esitetyt tulokset on saatu aina koko aineistosta, ellei toisin ole ilmoitettu. Tuloksia tarkasteltaessa on otettava huomioon, että 1.-luokkalaisten kirjoitelmien osuus koko aineistosta on hyvin pieni. Kun muiden luokkien kirjoitelmien yhteissanemäärät ovat yli 10 000 sanetta, 1.-luokkalaisten kirjoitelmien yhteissanemäärä jää selvästi alle 3 000 saneen. 1.-luokkalaisten kirjoitelmat vaikuttavat siis koko aineiston tuloksiin varsin vähän, minkä lisäksi on suhtauduttava varauksella siihenkin, mitä ne kertovat 1.-luokkalaisten sanavarastosta.

Aineisto koostuu neljästä eri korpuksesta, jotka on koodattu valmiiksi. Korpusten koodaus on pyritty yhdenmukaistamaan tätä tutkimusta varten, mutta joitakin eroavaisuuksia on saattanut jäädä. Heitot ovat kuitenkin marginaalisen pieniä, eivätkä vaikuta merkittävästi tuloksiin. Numerot on myös pidetty mukana aineistossa, ellei toisin ole ilmoitettu. Mikäli tarkasteltavana on esimerkiksi sata frekventeintä sanaa, sanalista ei ole katkaistu sadannen sanan kohdalta, vaan listalle on otettu mukaan kaikki ne lekseemit, joita on käytetty yhtä monta kertaa kuin sadanneksi yleisintä lekseemiä. Tästä syystä joillakin sadan yleisimmän lekseemin listoilla lekseemejä on enemmän kuin sata.

3.1.1 Frekvenssit ja lekseemien toisto

Saneiden ja lekseemien kokonaismäärät kertovat, paljonko saneita sekä erilaisia lekseemejä eri luokkien aineistossa on. Tässä alaluvussa tarkastellaan saneiden ja lekseemien määriä, sana-sane-suhdetta, lekseemien toistoa sekä sanaluokkien kumulatiivista frekvenssiä kaikkien vuosiluokkien aineistoissa. Lopuksi luodaan vielä silmäys siihen, kuinka lekseemien keskimääräinen lukumäärä per oppilas muuttuu luokka-asteelta toiselle siirryttäessä.



Kuvio 1. Saneiden ja lekseemien kokonaismäärät (sis. numerot).

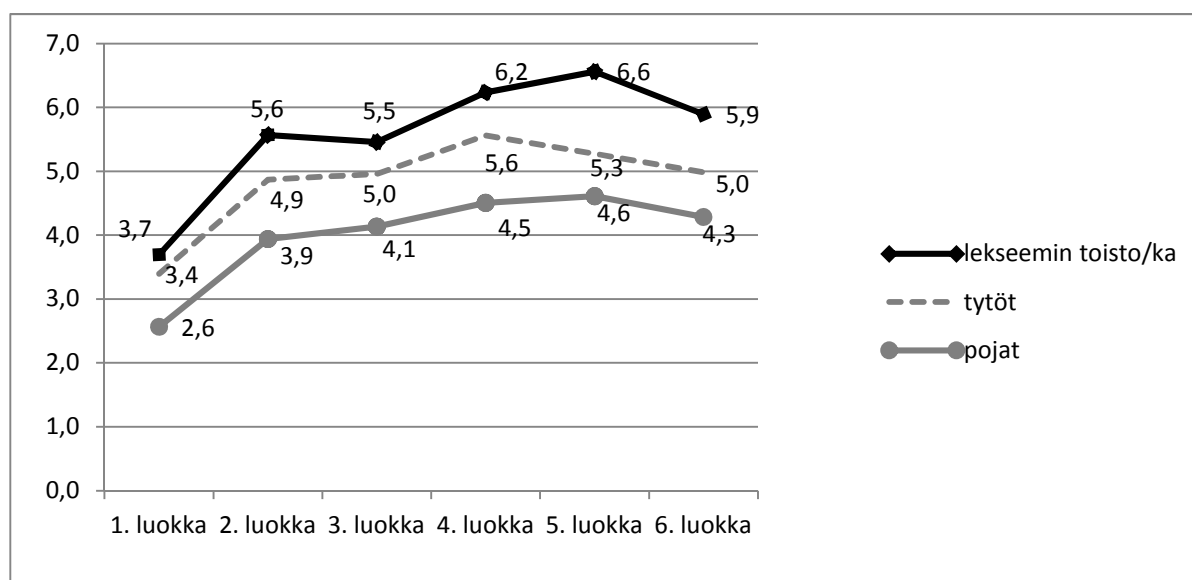
Kuviossa näkyy hyvin se, että lekseemien ja saneiden kuvaajat ovat varsin erimalliset. Se kertoo siitä, että lekseemien määrä ei kasva samassa suhteessa sanemäärän kanssa. Vaikka 2. luokalla saneita onkin 1. luokkaan verrattuna lähes nelinkertainen määrä, lekseemien määrä on vain vähän yli kaksinkertainen. 6.-luokkalailla on 1. luokkaan verrattuna 6,5-kertainen määrä saneita ja noin nelinkertainen määrä lekseemejä.

6. luokan aineistossa saneita on 2. luokkaan verrattuna noin 1,7-kertainen määrä ja lekseemejä noin 1,6-kertainen määrä. 3. luokkaan verrattuna 6.-luokkalaisten sanemäärä on 1,3-kertainen ja lekseemimäärä 1,2-kertainen, 5. luokkaan verrattuna 6.-luokkalaisten sanemäärä on 0,9-kertainen ja lekseemimäärä 0,95-kertainen. Tämän perusteella sanavarasto monipuolistuu lasten varttuessa kohtuullisen hitaasti – mutta silti varmasti.

Taulukko 2. Sana-sane-suhde luokittain.

| | 1. luokka | 2. luokka | 3. luokka | 4. luokka | 5. luokka | 6. luokka |
|-----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| sana-sane-suhde | 0,27 | 0,18 | 0,18 | 0,16 | 0,16 | 0,17 |

Sana-sane-suhde (taulukko 2) kuvaa samaa tilannetta kuin kuvio 1. Saneiden määrän ei tarvitse olla järin suuri, jotta tietynlainen perussanasto tulee esiin. Tämän jälkeen sanemäärien kasvu näyttää lisäävän ensisijaisesti toistoa ja vasta toissijaisesti lekseemien määrää. Sanemäärän kasvun toistoa lisäävä vaikutus näkyy myös seuraavassa kuviossa (kuvio 2). Koska koko aineistossa on yhteensä enemmän saneita kuin erikseen tyttöjen tai poikien aineistossa, myös toistoa on koko aineistossa enemmän kuin niissä.



Kuvio 2. Lekseemin keskimääräinen toisto koko aineistossa sekä erikseen tyttöjen ja poikien kirjoitelmissa.

Se, että luokittainen lekseemien toiston keskiarvo kasvaa sanemäärän kasvaessa kertoo siitä, että sanasto ei monipuolistu käsi kädessä sanemäärän kasvun kanssa. Samaan ilmiöön on törmätty aiemminkin (ks. esim. Saarela 1991: 39); lekseemien toisto lisääntyy sanemäärän kasvaessa, mikä voi antaa vaikutelman siitä, että pidempiä aineita kirjoittavien tyttöjen sanasto olisi lyhyempiä aineita kirjoittavien poikien sanastoa suppeampi. Tulosta ei kuitenkaan ole syytä tulkita niin eikä sana-sane-suhdetta tai toistoa ilmaisevia lukuja voi yksintein pitää hyvinä sanaston monipuolisuutta kuvaavina indekseinä.

Kumulatiivinen frekvenssi kuvaa, miten suuren osan tietty määrä yleisimpiä lekseemejä kattaa koko aineiston saneista. Koulukorpus sisältää noin 7 770 lekseemiä, joista jo sata yleisintä kattaa yli puolet saneista ja 500 yleisintä yli 75 % saneista. Yli tuhatta yleisintä sanaa ei kannata edes ottaa mukaan tarkasteluun, sillä tuhannen yleisimmän lekseemin joukko koostuu reilusti yli tuhannesta (tarkemmin 1062) lekseemistä, joita on käytetty vähintään kahdeksan kertaa. 2 000 yleisimmän lekseemin joukkoon pääsevät lekseemit, joita on aineistossa vähintään kolme kappaletta. Yhden esiintymän lekseemejä on noin 3 500.

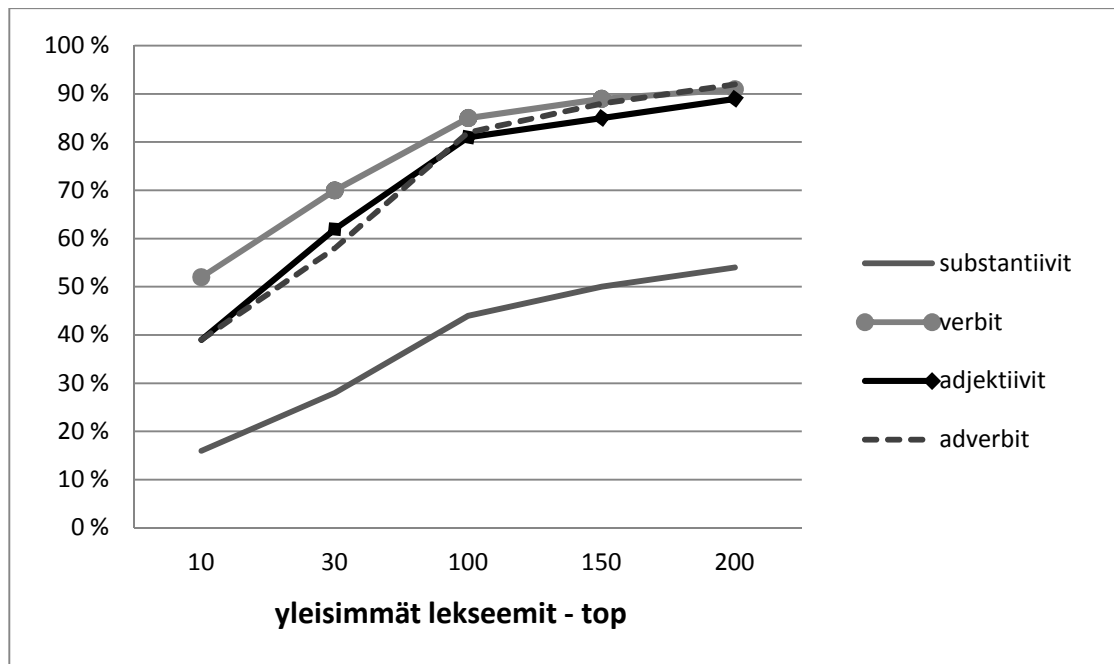
Vertailemalla Koulukorpus-aineistoa muihin aineistoihin saadaan käsitystä siitä, kuinka laajaa koulukirjoitelmien sanasto on. Alla olevassa taulukossa (taulukko 3) Koulukorpus-aineiston yleisimpien lekseemien kumulatiivista frekvenssiä verrataan peruskoulun 2. ja 3. luokan oppikirjoihin sekä CSC:n Suomen sanomalehtikielen taajuussanastoon.

Taulukko 3. Kumulatiivisen frekvenssin vertailua.

| yleisintä sanaa (kpl) | kattavuus Koulukorpuksessa (%) | kattavuus oppikirjatekstissä (%) (Jaakola 2004: 64) | kattavuus CSC:n sanomalehtikielen taajuussanastossa (%) |
|-----------------------|-----------------------------------|---|---|
| 20 | 33,8 | 21,0 | 15,7 |
| 50 | 45,7 | 29,4 | 21,0 |
| 100 | 55,3 | 37,4 | 26,1 |
| 500 | 77,1 | 61,3 | 42,7 |
| 1000 | 85,1 | 71,7 | 50,2 |

Kumulatiivinen frekvenssi vaihtelee paljon aineistosta riippuen. Siihen vaikuttavat ainakin sekä aineiston koko että aineiston alkuperä. Koulukirjoitelmissa, joiden aihepiiri on sama, yhteistä sanastoa esiintyy paljon jo pelkästään aihepiirin vuoksi. Ei ole lainkaan ihmeellistä, että eri alojen oppikirjoissa, vaikka ne onkin suunnattu 2. ja 3. -luokkalaisille, sanasto on eriytyneempää ja että sanomalehtiteksteissä, joita on koottu kaikenlaisten otsikoiden alta, sanaston frekvenssi kumuloituu vieläkin hitaammin.

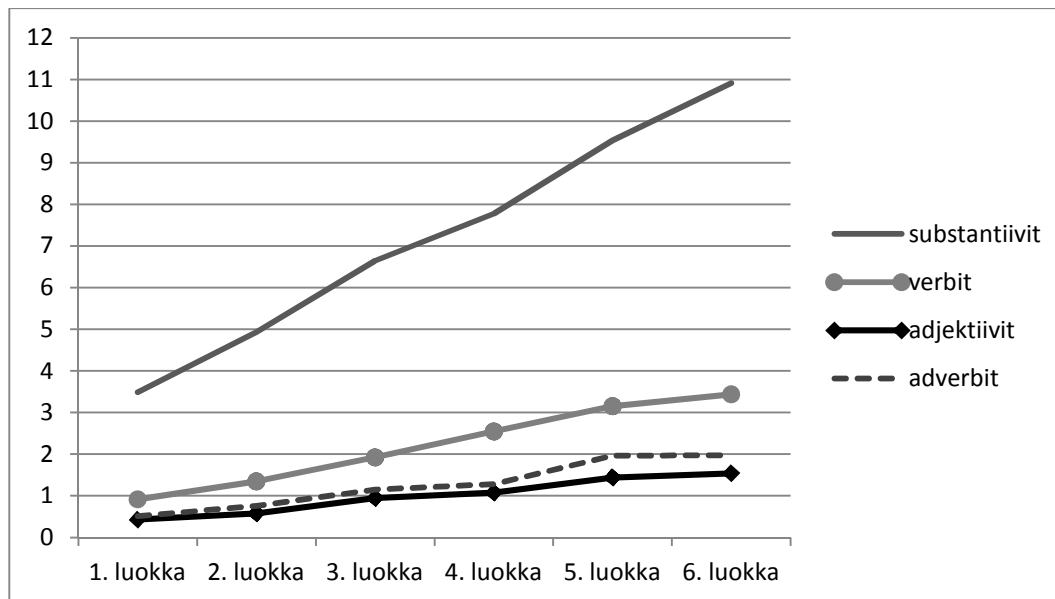
Seuraavaksi tarkastellaan vielä 200 yleisimmän lekseemien kumulatiivista frekvenssiä eri sanaluokkien osalta koko Koulukorpus-aineistossa.



Kuvio 3. Sanaluokkien kumulatiivinen frekvenssi koko aineistossa.

Kumulatiivinen frekvenssi kuvaa siis sitä, kuinka suuren osan tietty osa lekseemeistä kattaa, esimerkiksi kuinka suuren osan kymmenen yleisimmän verbilekseemin osuus kattaa koko aineiston verbisaneista. Substantiivit eroavat huomattavasti adjektiiveista, verbeistä ja adverbista siinä, että vasta 200 yleisintä lekseemiä kattaa yli 50 % kaikista substantiivisaneista. Muut sanaluokat ovat ylittäneet tämän jo 30 yleisimmän lekseemin sisällä. Tulos kertoo siitä, että substantiivilekseemejä käytetään monipuolisimmin ja niitä myös tunnetaan eniten. Sen sijaan kaikenlainen kuvailu ja erityisesti tekojen esittäminen tehdään enimmäkseen melko suppean perussanaston avulla. Substantiivilekseemien käyttö olisi todennäköisesti vieläkin eriytyneempää, mikäli kirjoitelmien aihe olisi täysin vapaa.

Lopuksi tarkastellaan vielä sitä, kuinka monta substantiivi-, adjektiivi, verbi- ja adverbilekseemiä keskimäärin per oppilas löytyy eri luokkien aineistosta.



Kuvio 4. Lekseemien lukumäärä yhtä oppilasta kohti keskimäärin.

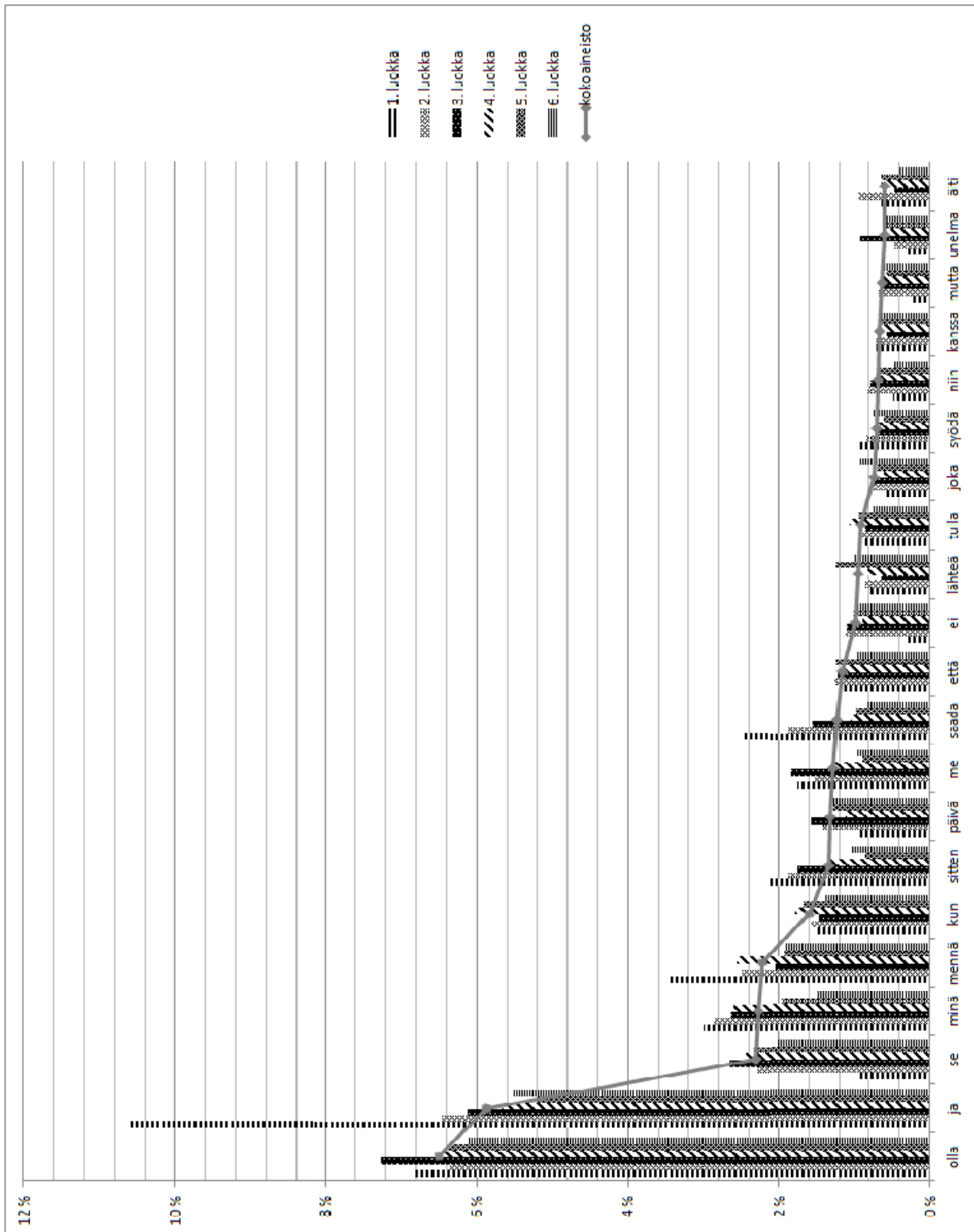
Kuvion 4 perusteella leksiemien lukumäärä oppilasta kohti lisääntyy selvästi oppilaiden varttuessa. Substantiivien kohdalla leksiemimäärä kasvaa hyvin tasaisesti 4. luokan kohdalla näkyvää pientä notkahdusta lukuun ottamatta. Myös verbien osalta leksiemimäärän kasvu on hyvin tasaista aina 5. luokkaan saakka, minkä jälkeen kasvu näyttää hidastuvan. Samankaltainen kasvun hidastuminen näkyy myös adjektiivien ja adverbien kohdalla, joiden määrissä näkyy lisäksi myös notkahdus 4. luokan kohdalla.

Miksi 4. luokan kohdalla on notkahdus? Kysymykseen ei ole helppo vastata. 3. luokalla kirjoittajien joukossa on kielikylpyluokka, jonka kirjoitustaidot eivät välttämättä ole aivan tavallisten koululaisten tasolla, mutta 4. luokalla tämän eron pitäisi kuitenkin jo tasoittua. Toisaalta 3.-luokkalaisten joukossa on kielikoululaisia, jotka taas ovat keskimäärin taitavampia kirjoittajia kuin tavalliset koululaiset. Saattaa olla, että 3. luokan aineiston heterogeenisyys, toisin sanoen kirjoittajien erilainen koulutausta, rikastaa leksikkoa jonkin verran tavallisesta poikkeavasti, mistä syystä 4.-luokkalaisten kohdalla käyrässä tapahtuu notkahdus. Joka tapauksessa leksiemien määrä kasvaa jatkuvasti riippumatta siitä, montako kirjoittajaa luokilla on, mikä osoittaa sen, että ainakin jokaisen vuosiluokan yhteinen sanasto karttuu jatkuvasti, vaikka yksilöiden kohdalla tilanne voi olla hyvin epätasainen.

3.1.2 Yleisimmät lekseemit

Tässä alaluvussa tarkastellaan koko Koulukorpus-aineiston yleisimpiä lekseemejä sekä vertaillaan 2., 4. ja 6. luokan kirjoitelmien sadan yleisimmän lekseemin sanaluokkia.

Suurin osa aineiston yleisimmistä lekseemeistä on perussanastoa tai aineen aiheeseen kiinteästi liittyvää sanastoa, jota on käytetty kaikkien vuosiluokkien kirjoitelmissa (kuvio 5 ja liite 2). Kuviossa viiva kuvaa kunkin lekseemin yleisyyttä koko aineistossa ja palkit lekseemien frekvenssiä luokittain.



Kuvio 5. Koko aineiston 21 yleisintä lekseemiä ja niiden osuudet luokittain.

Erityisesti 1.-luokkalaisilla monet yleisimmistä lekseemeistä ovat hyvin frekventtejä, mikä kertoo sekä sanaston yksipuolisuudesta että keskittymisestä tiettyihin peruslekseemeihin. Vanhempien oppilaiden, ja erityisesti 5.- ja 6.-luokkalaisten aineissa, yleisimmät lekseemit eivät saa enää yhtä suuria osuuksia, joskin 3. luokka tekee joidenkin lekseemien kohdalla poikkeuksen, mikä saattaa johtua 3. luokan muita luokkia heterogeenisemmasta oppilasaineuksesta.

Koulukorpus-aineiston yleisimmät lekseemit ovat perussanoja, joista joidenkin yleisyys perustuu selvästi joko kirjoittajien ikään, aineen aiheeseen tai kirjoitelmien tekstilajiin. Tällaisia lekseemejä ovat todennäköisesti ainakin *minä*, *sitten*, *syödä*, *kanssa*, *unelma* ja *äiti*. *kanssa* on hyvin frekventti siksi, että unelmien päivää vietetään harvoin yksin. *sitten* on tavallinen tapa osoittaa siirtymää tapahtumissa ja *syödä*-verbin yleisyys seuraa siitä, että herkkujen syöminen näyttää olevan unelmien päivänä hyvin tavallista. 21 yleisimmästä lekseemistä *olla*, *ja*, *se*, *saada*, *että*, *ei* ja *joka* ovat CSC:n taajuussanaston kymmenen yleisimmän lekseemin joukossa ja *mennä*, *kun*, *päivä*, *me*, *tulla*, *niin* sekä *mutta* sadan yleisimmän lekseemin joukossa. Niinkin tavalliselta kuulostavat *minä* tai *sitten*, sekä aineen aiheen vuoksi frekventti *unelma* eivät mahdu CSC:n taajuussanastossa tuhannen yleisimmän lekseemin joukkoon.

Joka tapauksessa Koulukorpus-aineiston leksikon sanat ovat lyhyitä aikuiskielen leksikkoon verrattuna. Aiemmissa tutkimuksissa onkin jo havaittu, että koko Koulukorpus-aineiston sanapituudet keskittyvät 4–9 merkkiin ja että saneiden mediaanipituus on kuusi merkkiä ja lekseemien viisi merkkiä (Pajunen 2012: 14–16). Vaikka kielen yleisimmät lekseemit ovat myös aikuisten käyttämässä sanastossa lyhyehköjä, jo sadan yleisimmän lekseemin vertailu Koulukorpuksen ja CSC:n taajuussanaston välillä osoittaa pienen eron lekseemipituuksissa: Kun Koulukorpusaineiston sadan yleisimmän lekseemin pituus on 4,7 merkkiä, CSC:n sadan yleisimmän lekseemin pituus on 4,8 merkkiä. Oppikirjasanaston sadan yleisimmän lekseemin keskipituus on 5,3 merkkiä, mikä ehkä tavallaan kuvaa oppikirjasanaston erityisluonnetta: Jo sadan yleisimmän lekseemin joukkoon mahtuu pitkäköjiä eri oppiaineille tyypillisiä erityissanoja, joiden rooli lasten kielenkäytössä on varsin minimaalinen.

Vertailemalla Koulukorpus-aineiston sataa yleisintä lekseemiä Opettajan sanastokirjan, CSC:n sanomalehtikielen taajuussanastoon sekä 2.- ja 3.-luokkalaisten oppikirjojen sanastoon voidaan hahmottaa, mitä aineiston lekseemejä on käytetty ennen kaikkea Unelmien päivä - aiheen vuoksi ja mitä lekseemejä voi pitää kielen perussanastoon kuuluvana. Opettajan

sanastokirjan sadan yleisimmän lekseemin listalla on 61, CSC:n taajuussanaston listalla 50 ja oppikirjojen listalla 32 samaa lekseemiä kuin Koulukorpus-aineiston sadan yleisimmän lekseemin listalla. Lekseemejä, joita ei Koulukorpus-aineiston lisäksi löydy minkään verrokkiaineiston sadan yleisimmän sanan listalta, on 28. Ne ovat *aamupala, hakea, herätä, heti, hevonen, hieno, hotelli, ihana, ilta, jokin, karkki, kaveri, kiva, koska, kysyä, leikkiä, loppu, nukkua, paras, pelata, ruoka, seuraava, toivoa, tosi, uida, unelma, voittaa* ja *ystävä*. Koska Koulukorpuksen aineet käsittelevät unelmien päivää, päivän ja tapahtumien kulkua jäsentävät lekseemit kuten *herätä, nukkua, aamupala* ja *ilta* sekä *seuraava* ja *loppu* ovat yleisimpien joukossa. Lasten kiinnostuksenkohteita ja harrastuksia edustavat lekseemit *hevonen, karkki, kaveri, leikkiä, pelata, ruoka, uida, voittaa* ja *ystävä*. Frekventtejä kirjoitelmille annetun aiheen vuoksi ovat *unelma, hotelli* ja todennäköisesti myös *toivoa*. Lasten tavalliseen kielenkäyttöön puheessakin kuuluvat *kiva* ja vahvikesana *tosi*, joita taas ei niinkään lehti- tai oppikirjatekstissä esiinny. 1960-luvun koululaistenkin eniten käyttämien lekseemien listalla *kiva* on vasta sijalla 432 (Opettajan sanastokäsikirja 1970: 92).

Seuraavaksi tarkastellaan 2.-, 4.- ja 6.-luokan kirjoitelmien sadan yleisimmän lekseemin sanaluokkia (taulukko 4).

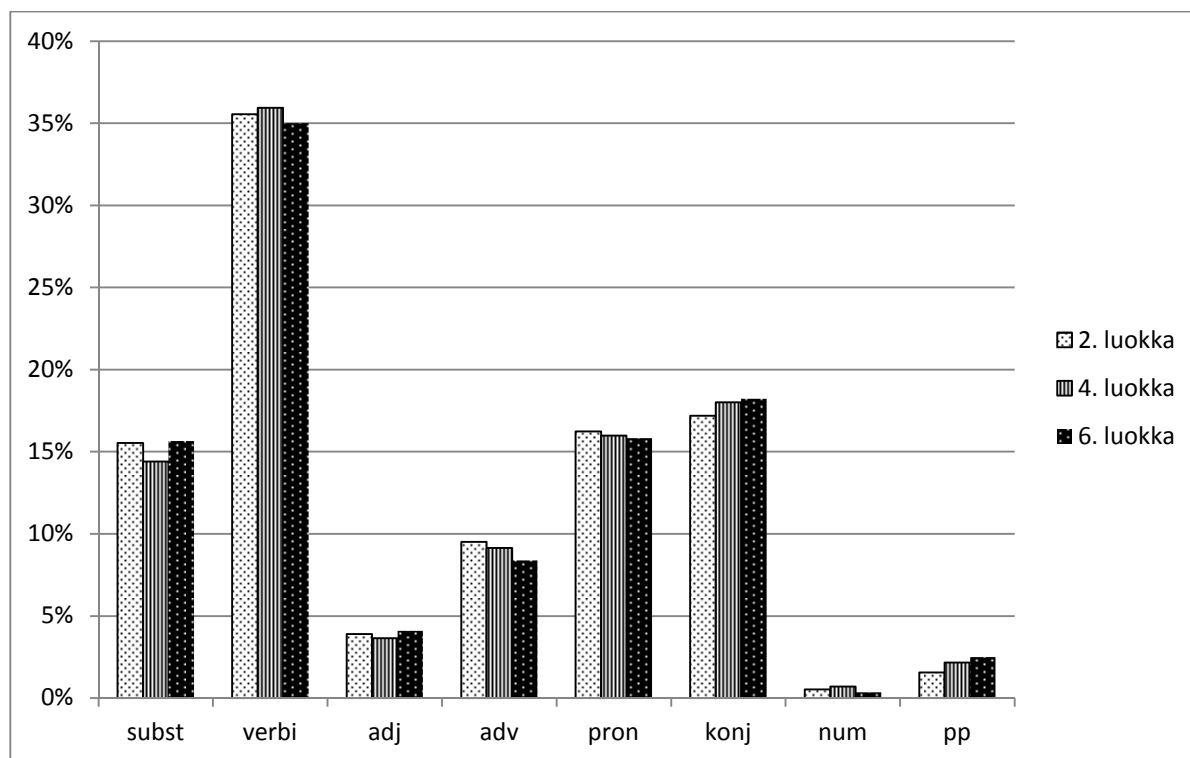
Taulukko 4. 2.-, 4.- ja 6.-luokan sadan yleisimmän lekseemin (liite 7) sanaluokat.

| | 2. luokka % (lkm) | 4. luokka % (lkm) | 6. luokka % (lkm) |
|---------------|-------------------|-------------------|-------------------|
| substantiivit | 26,7 (27) | 25,5 (27) | 25,5 (26) |
| verbit | 28,7 (29) | 27,4 (29) | 27,5 (28) |
| adjektiivit | 8,9 (9) | 10,4 (11) | 9,8 (10) |
| adverbit | 14,9 (15) | 14,2 (15) | 12,7 (13) |
| pronominit | 9,9 (10) | 11,3 (12) | 12,7 (13) |
| konjunktiot | 6,9 (7) | 7,5 (8) | 8,8 (9) |
| numeraalit | 2,0 (2) | 1,9 (2) | 1,0 (1) |
| adpositiot | 2,0 (2) | 1,9 (2) | 2,0 (2) |
| kaikki | 100 (101) | 100 (106) | 100 (102) |

2.-luokan sadan yleisimmän lekseemin joukko koostuu lekseemeistä, jotka esiintyvät aineistossa 16–676 kertaa. 4.-luokkalaisten yleisimpien lekseemien joukkoon pääsivät lekseemit, jotka esiintyvät luokan kirjoitelmissa 27–1 222 ja 6.-luokkalaisten yleisimpien lekseemien joukko koostuu 26–1 058 kertaa esiintyneistä lekseemeistä. 2.-luokkalaisten sadan yleisimmän lekseemin joukkoon mahtuu enemmän verbejä, substantiiveja, adverbejä ja numeraaleja, 4.- ja 6.-luokkalaiset sen sijaan käyttävät runsaammin adjektiiveja, pronomineja

ja erilaisia konjunktioita. Lukumääräisesti erot ovat varsin pieniä, ja ne saadaankin näkymään selvemmin prosentuaalisen tarkastelun avulla.

Seuraavaksi tarkastellaan sadan yleisimmän lekseemin sanaluokkien osuuksia saneista 2., 4. ja 6. luokkien kirjoitelmissa (kuvio 6).



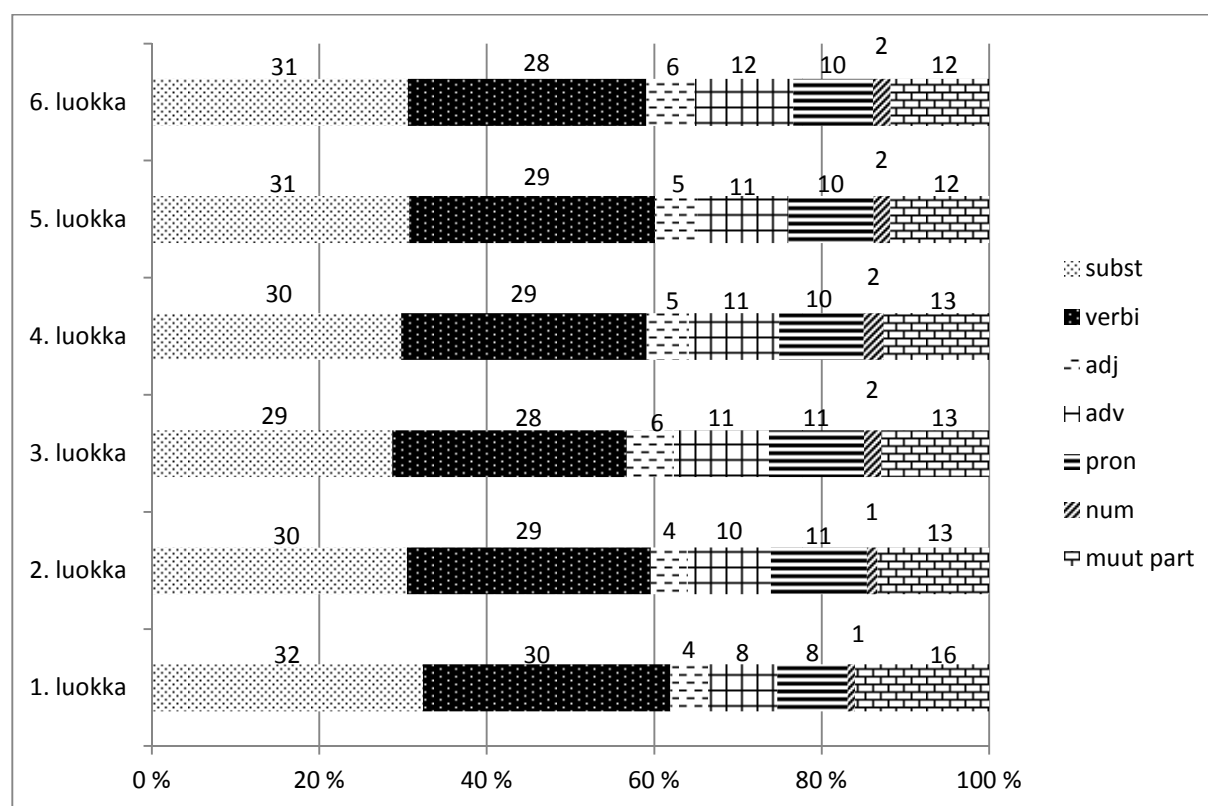
Kuvio 6. Sadan yleisimmän lekseemin sanaluokkien osuudet saneista.

Sanaluokkien osuudet sadan eniten käytetyn lekseemin joukosta eivät eroa paljon luokka-asteittain. Joitakin pieniä eroavaisuuksia kuitenkin voidaan osoittaa. Substantiivien osuus on pienimmillään 4. luokalla ja verbien 6. luokalla. Adverbien ja pronomien osuudet pienyvät ja konjunktioiden ja adpositioiden osuudet kasvavat hieman luokka-asteelta ylemmälle siirryttäessä. Substantiivien, verbien, adjektiivien ja numeraalien osuudet taas eivät noudata selkeää linjaa. Verbien suuri osuus johtuu todennäköisimmin siitä, että kirjoitelmissa käytetään paljon samoja perusverbejä.

3.2 Sanaluokkien osuudet

Tässä aluvussa keskitytään saneiden ja lekseemien sanaluokkien osuuksiin. Sanaluokat on jaettu adjektiiveihin, numeraaleihin, pronomineihin, substantiiveihin, verbeihin, adverbeihin ja muihin partikkeleihin, joka sisältää interjektioita, konjunktioita ja adpositioita.

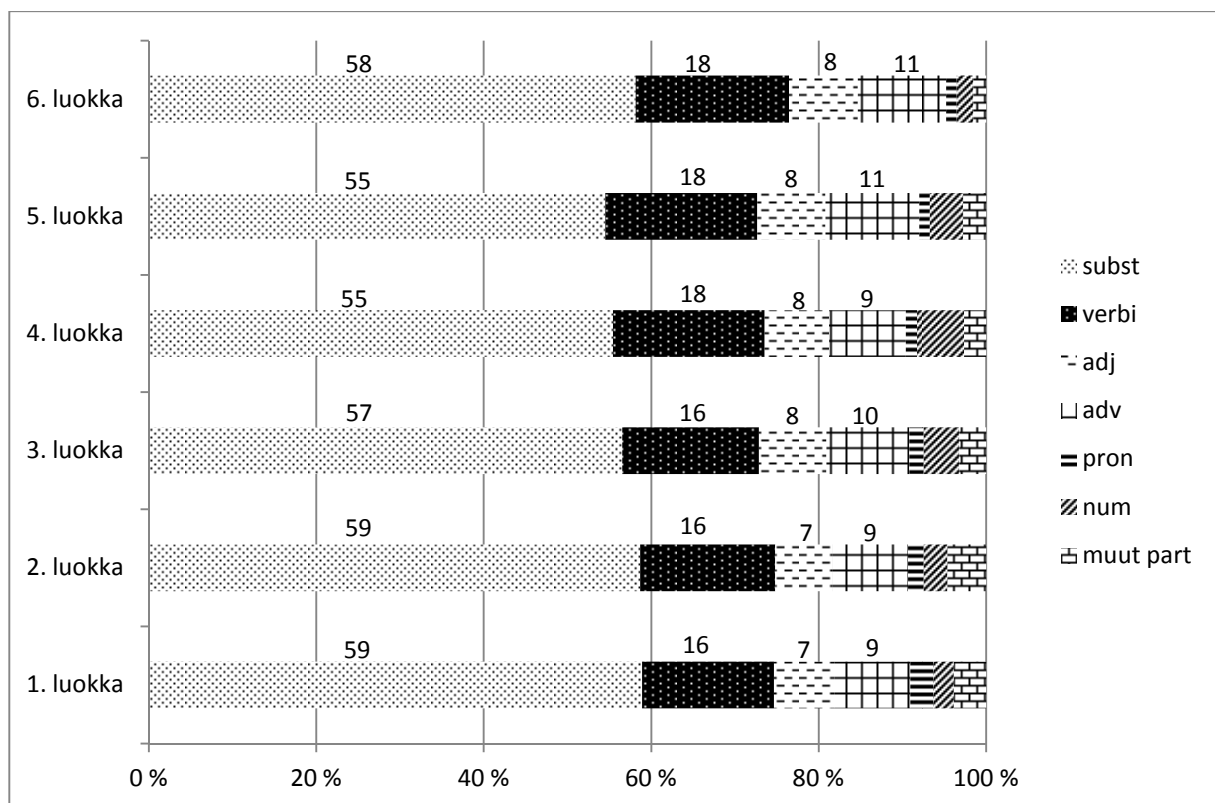
Kovin suuria luokka-asteiden välisiä eroja sanaluokkien osuuksissa ei ole (kuvio 7). 1.-luokkalaisten käyttävät eniten muita partikkeleita, 6.-luokkalaisten taas eniten adverbeja, joskin erot muihin luokkiin ovat useimmiten vain yhden tai kahden prosenttiyksikön luokkaa. Adjektiivien ja adverbien osuus kirjoitelmissa kasvaa selvästi luokka-asteelta ylöspäin siirryttäessä, kun taas muiden partikkelien käyttö vähenee. Substantiivien ja verbien osuus vaihtelee. Yleisesti ottaen sanaluokkien osuudet eivät lisäänty tai vähene lineaarisesti luokka-asteelta toiselle siirryttäessä, vaan ne vaihtelevat epälineaarisesti. Siksi sanaston yleistä kehitystä arvioitaessa onkin parempi kiinnittää huomiota yleiseen trendiin kuin kahden peräkkäisen luokka-asteen välisiin eroihin.



Kuvio 7. Saneiden sanaluokkien osuudet 1.–6.-luokkalaisten kirjoitelmissa.

Osittain vaihtelu eri luokkien välillä on niin pientä, että prosenttiyksikön, parin eroihin ei ole syytä kiinnittää suuremmin huomiota. 3. ja 4. luokan tulosten osalta on muistettava, että mukana on kielikylpyluokka, jonka oppilaiden sanasto saattaa poiketa jonkin verran tavallisten luokkien oppilaiden sanastosta. Kielikylpyoppilaiden kirjoitelmien sanemäärä kattaa 3.-luokkalaisten aineistosta vajaat 20 % ja 4.-luokkalaisten aineistosta noin 23 %. Lisäksi 3. luokan aineiston saneista noin 23 % on kielikoulun oppilailta. Kielikouluryhmän erityispiirteenä on kirjoitelmien keskimääräistä korkeampi taso muuhun 3. luokan aineistoon verrattuna.

Kun tarkastelun kohteena ovat lekseemien sanaluokkien osuudet (kuvio 8), erityisesti substantiivien ja verbien osuudet muuttuvat huomattavasti. Avoimet sanaluokat, eli substantiivit, verbit, adjektiivit ja adverbit, kattavat lekseemeistä huomattavasti suuremman osuuden kuin saneista.



Kuvio 8. Eri lekseemien sanaluokkien osuudet 1.–6.-luokkalaisten kirjoitelmissa.

Substantiivit kattavat kaikilla luokkatasoilla yli puolet kaikista lekseemeistä (kuvio 8). Verbien osuus on alle 20 %, ja se kasvaa hieman ylemmillä luokka-asteilla. Samoin kasvavat myös adjektiivien ja adverbiin osuudet. Ylemmillä luokilla pienemmän osuuden saavat sen

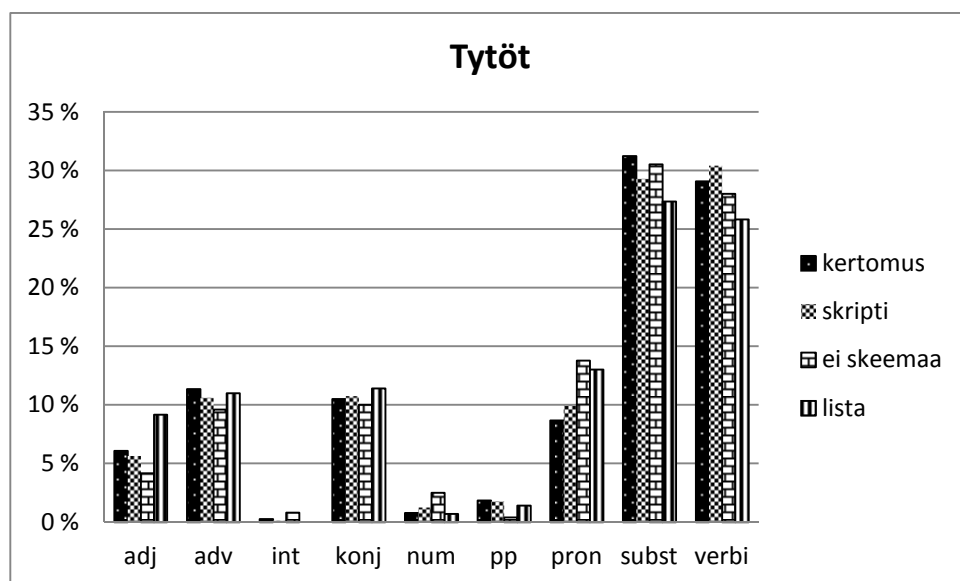
sijaan pronominit ja muut partikkelit, joihin kuuluvat interjektit, konjunktiot ja adpositiot. Tämä kuvastaa sitä, kuinka ylemmillä luokilla lekseemien määrä kasvaa, ja erityisesti verbi-, adjektiivi- ja adverbisanasto rikastuvat. Substantiivilekseemien verrattain suuri osuus 6. luokan aineistossa saattaa kertoa siitä, että substantiivilekseemien määrä on kasvanut hyvin paljon 5. luokkaan verrattuna, mutta syitä voi olla muitakin. Laadullinen tarkastelu tuo esiin sen, että vaikka määrällisesti jonkin sanaluokkien osuuksien vaihtelu on pientä, esimerkiksi adjektiivisanasto kompleksistuu ylemmillä vuosiluokilla (Kuisma 2012: 38). Vaikka ylemmillä vuosiluokilla käytetään alaluokkia monipuolisemmin myös pronomineja ja konjunktioita, näiden sanaluokkien osuus ei voi kasvaa loputtomiin, koska ne ovat suljettuja, mikä on syynä pronomien sekä yhtenä syynä myös muiden partikkelien osuuden pienenemisessä. Muihin partikkeleihin olen konjunktioiden lisäksi laittanut myös adpositiot ja interjektit, joista adpositiolekseemien osuuden vähenemiseen vaikuttaa ainakin se, että ne ovat suljettu luokka, mutta mahdollisesti sekin, että adposition käyttö voidaan korvata sijamuodolla, esimerkiksi *pöydän päälle* pro *pöydälle*. Seuraava 2.-luokkalaisten tytön kirjoitelmasta löytyvä esimerkki valottaa mahdollista, ehkä enemmän nuorilla kirjoittajilla esiintyvää adpositioiden ”liikakäyttöä”:

- 1) -- siellä olisi poreallas ja liukumäki ja tartsanköysi ja patjoja kellumassa veden päällä --. (kaupunki 2T5)

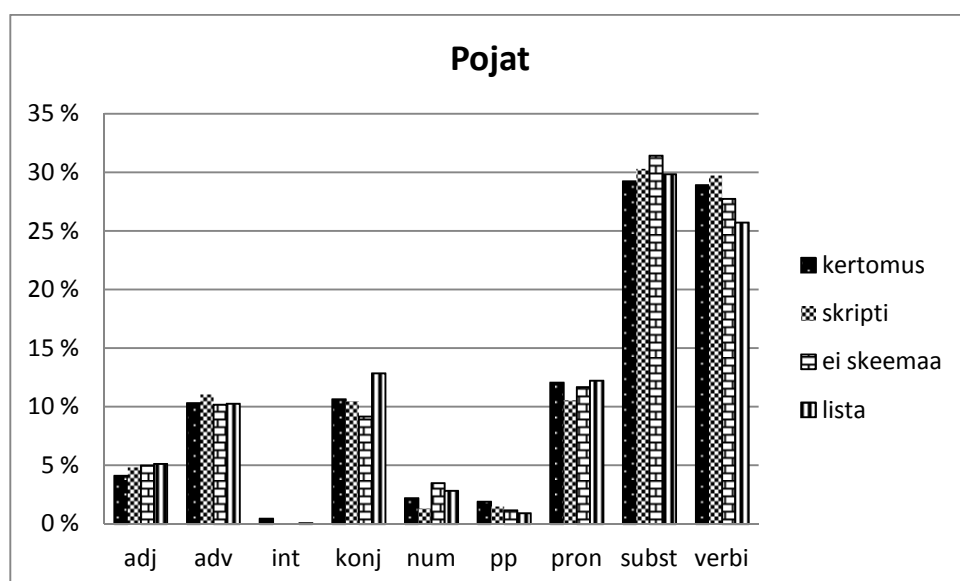
Interjektioihin ei tässä tutkimuksessa tarkemmin paneuduta. Periaatteessa interjektit ovat avoin luokka, mutta kun oikeinkirjoituksen ja kerronnan konventiot vahvistuvat ja epäsuoraesitys sekä referointi lisääntyvät kirjoittajien tekstissä, interjektien käyttö epäilemättä vähenee suoran dialoginkin vähentyessä, ja mahdolliset interjektit alkavat olla vähemmän luovia, kirjoitetussa yleiskielessä tavallisemmin tavattavia. Esimerkiksi 2.-luokkalaisten kirjoitelmissa esiintyy 31 eri interjektiolekseemiä, mutta 6.-luokkalaisten eri interjektiolekseemejä on enää 20, vaikka 6.-luokkalaisten sanavarasto onkin 2.-luokkalaisiin verrattuna huomattavasti runsaampi. Molempien luokkien kirjoitelmissa vain viisi interjektiolekseemiä esiintyy useammin kuin kerran, ja eniten käytetty interjektio on *no* (esimerkit 2 ja 3).

- 2) Tiedätkö, minkä näköinen saksanpaimenkoira on? No minäpä kerron. (kaupunki 2T47)
- 3) No kun vihdoinkin saimme pelin päätökseen ja voittajaksi olin tullut minä, me istahdimme sohvalle odottamaan joulupukkia. (taajama 6T43)

Seuraavaksi tarkastellaan sanaluokkien osuutta saneista erikseen tytöillä ja pojilla kirjoitelmien tason mukaan (kuviot 9 ja 10).



Kuvio 9. Saneiden sanaluokkien osuudet tyttöjen eritasoisissa kirjoitelmissa.



Kuvio 10. Saneiden sanaluokkien osuudet poikien eritasoisissa kirjoitelmissa.

Kovin selkeitä yhteyksiä sanaluokkien osuuksien ja kirjoitelmien tason välillä ei ole. Adjektiiveja käytetään melko tasaisesti kaikentasoisissa kirjoitelmissa lukuun ottamatta tyttöjen lista-tasoisia kirjoitelmia, joissa niitä käytetään huomattavasti enemmän kuin muissa kirjoitelmissa. Vain ei skeemaa -tasoisissa kirjoitelmissa pojat käyttävät adjektiiveja enemmän kuin tytöt. Adverbien käytössä ei ole suuren suuria eroja. Eniten niitä käytetään tyttöjen kertomus-tasoisissa kirjoitelmissa. Konjunktoiden osuus on suurin lista-tasoisissa

kirjoitelmissa sekä pojilla että tytöillä, mikä saattaa johtua luettelomaisesta kerronnasta ja siihen liittyvästä runsaasta *ja*-sanana käytöstä. Pojat käyttävät numeraaleja tyttöjä runsaammin lähes kaikentasoisissa kirjoitelmissa, ja vain skripti-tasoisissa kirjoitelmissa tytöt ja pojat käyttävät numeraaleja yhtä paljon.. Runsainta numeraalien käyttö on sekä tytöillä että pojilla ei skeemaa -tasoisissa kirjoitelmissa. Pronominien osuus näyttäisi tytöillä jonkin verran lisääntyvän, kun kirjoitelmien taso laskee, kun taas pojilla pronominien käyttö on runsainta sekä parhaissa että heikoimmissa kirjoitelmissa. Tason laskiessa kasvaa myös substantiivien osuus poikien kirjoitelmissa. Sen sijaan tyttöjen kirjoitelmissa substantiivien osuus ja verbien osuus sekä tyttöjen että poikien kirjoitelmissa pienenee kirjoitelmien tason heiketessä.

Vaikka Jaakolan aineistoa, Suomen sanomalehtikielen taajuussanastoa ja Koulukorpusta ei lekseemitasolla juuri kannata vertailla keskenään, sanaluokkien jakaumaa voidaan vielä tarkastella (taulukko 5).

Taulukko 5. Sadan yleisimmän lekseemin sanaluokat.

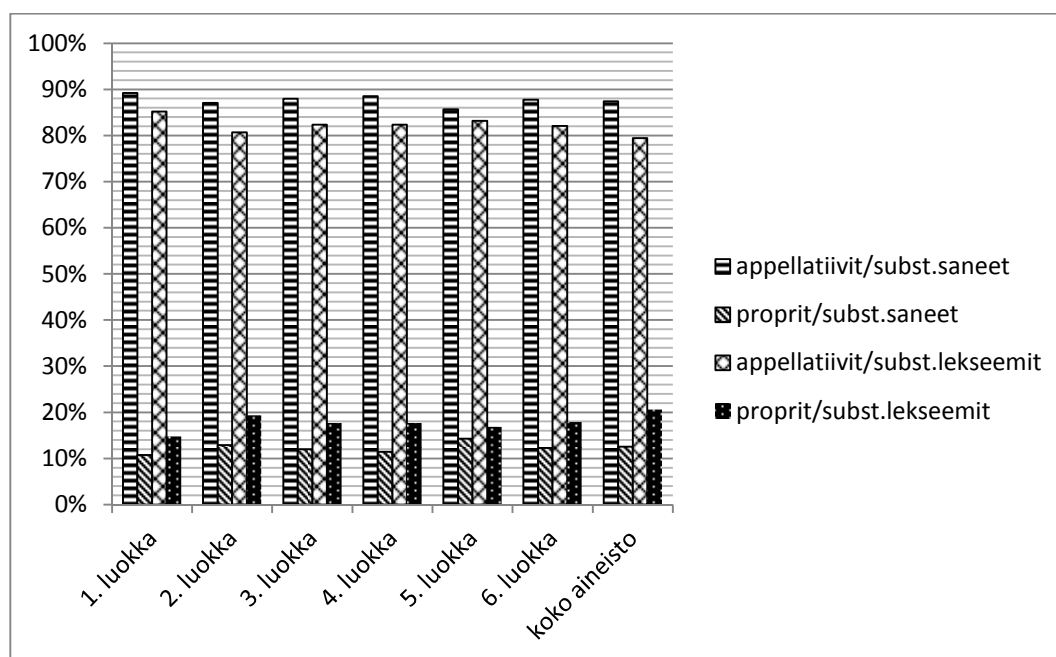
| sanaluokka | Koulukorpus (kpl) | oppikirjat (kpl) Jaakola 2004:liite 1 ¹ | taajuussanasto (kpl) |
|---------------|---------------------|---|----------------------|
| substantiivit | 24 (sis. 0 propria) | 37 (sis. 2 propria) | 22 (sis. 3 propria) |
| adjektiivit | 12 | 14 | 8 |
| verbit | 28 | 25 | 23 |
| adverbit | 9 | 5 | 10 |
| pronominit | 13 | 9 | 15 |
| konjunktiot | 10 | 6 | 10 |
| adpositiot | 2 | 0 | 3 |
| numeraalit | 2 | 4 | 6 |
| lyhenteet | 0 | 0 | 3 |

Taulukosta näkee, että tekstilajilla on suuri vaikutus sanaluokkien osuuksiin. Oppikirjasanasto poikkeaa Koulukorpus-aineiston sanastosta huomattavan paljon lähes jokaisen sanaluokan osalta. Koulukorpus eroaa molemmista vertailuaineistoista selvimmin verbien osalta. Mielenkiintoista on, että Koulukorpus-aineisto ja sanomalehtikielen taajuussanasto näyttävät usein olevan enemmän samoilla linjoilla oppikirjatekstin erotessa molemmista merkittävästi. Vain adjektiivien, verbien ja numeraalien käyttö on Koulukorpuksen osalta lähempänä oppikirjatekstien kuin sanomalehtikielen lukemia. Ilmeisesti sanomalehtitekstit ovat tekstilajina kuitenkin lähempänä koulukirjoitelmia kuin oppikirjateksti.

¹ Jaakolan sanaluokkakajoa on yhtenäistetty tämän tutkimuksen luokittelun kanssa.

3.2.1 Substantiivit

Substantiivit eli nimisanat jaotellaan yleisnimiin eli appellatiiveihin ja erisnimiin eli propriihin. Luokkien välinen vertailu (kuvio 11) osoittaa, että appellatiivien ja proprien suhde vaihtelee luokittain hivenen, mutta niiden välisessä suhteessa ei tapahdu selkeää muutosta lasten varttuessa.



Kuvio 11. Appellatiivien ja proprien osuudet substantiivisaneista ja -lekseemeistä.

2.-luokkalaisten 123 frekventeimmän substantiivin (vähintään 5 kappaletta) joukkoon mahtuu 16 propria, joista eniten käytetty, *Liisa*, oli 15:nneksi käytetyin. Loput propriit olivat *Särkänniemi*, *Jukka*, *Eppu*, *Lappi*, *Suomi*, *Aku*, *Helsinki*, *Intia*, *Pena*, *Playstation*, *Pokemon*, *Ruotsi*, *Thanh*, *Veeti* ja *Vilma*. Useat propriit esiintyvät vain yhdessä kirjoitelmassa, ja niiden runsasfrekventtisyys johtuu siitä, että kirjoittaja ei välttele toistoa esimerkiksi pronomineja käyttämällä (esimerkit 4 ja 5).

- 4) Liisa meni syömään ja kun liisa söi niin Liisa menee ulos ja sit Liisa menee tansitunille ja kun liisa tulee hän menee kaverille ja kun liisa tulee han kiipee puuhun ja Liisa putoo ja tulee haava polveen -- (kaupunki 2T33)
- 5) Muutaman sekunnin päästä peikko oli Intiassa. Intia oli hieno maa. Peikko asui kolme vuotta Intiassa ja palasi. (taajama 2T111)

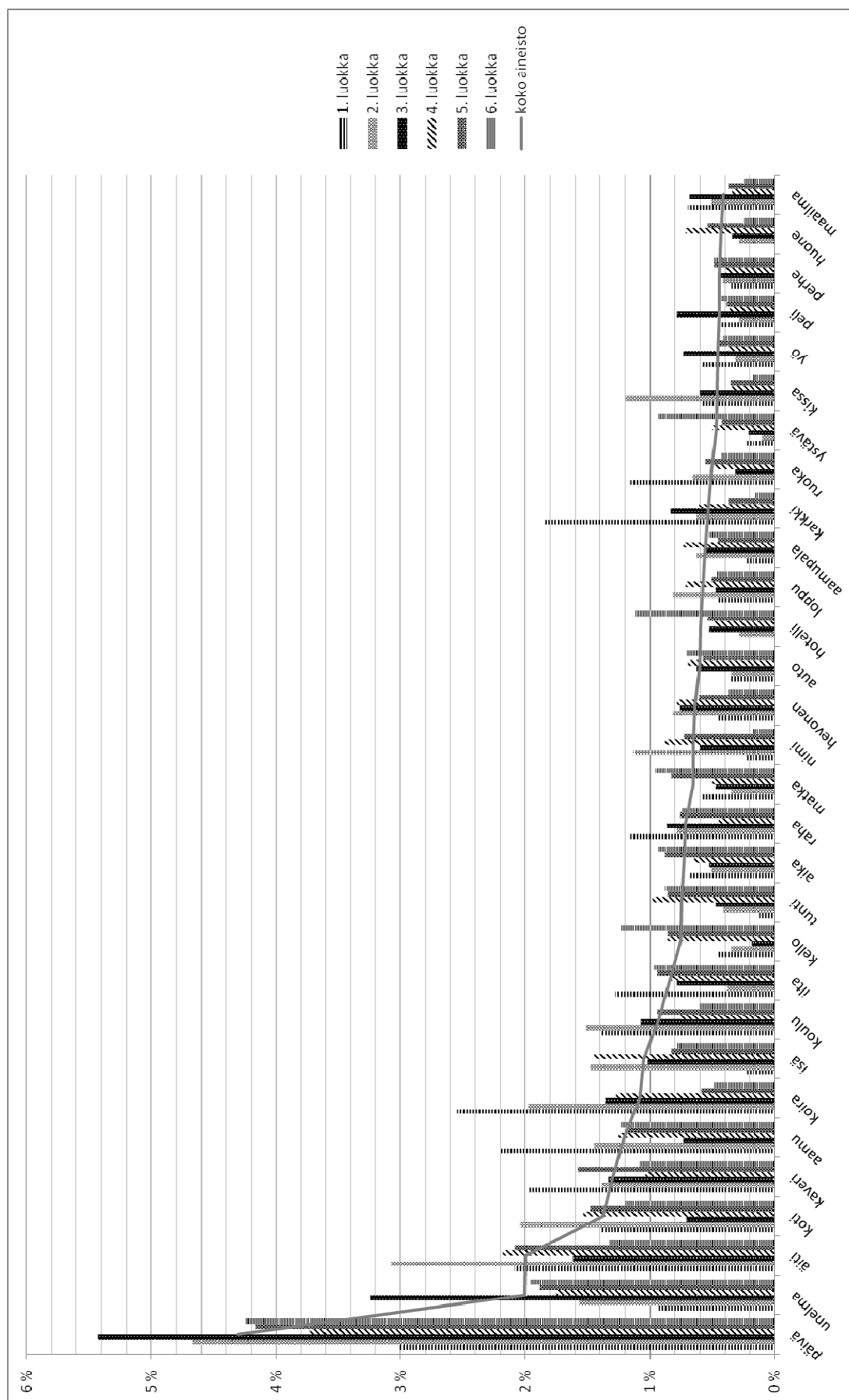
Esimerkiksi Liisasta kertovassa tarinassa (esimerkki 4) on yhteensä 93 sanetta, ja *Liisa* on toistettu 13 kertaa.

4.-luokkalaisten 101 frekventeimmän substantiivin (vähintään 10 kappaletta) joukkoon mahtuu enää vain kuusi propria, joista eniten käytetty, *Manu*, löytyy yleisimpien substantiivien listasta sijalta 65. Muut yleisimmät propriit ovat *Helsinki*, *Elli*, *Suomi*, *Ruotsi* ja *Särkänniemi*. Myös 6.-luokkalaisten 103 frekventeimmän substantiivin (vähintään 10 kappaletta) joukkoon mahtuu kuusi propria. Niistä eniten käytetty, *Helsinki*, löytyy sijalta 21, muut listan propriit ovat *Suomi*, *Särkänniemi*, *Amerikka*, *Stiina* ja *Paula*. *Paulaa* on käyttänyt kaksi kirjoittajaa, *Stiinaa* vain yksi. *Stiinan* suureen frekvenssiin (11 kappaletta) on syynä se, että kyseinen kertomus on kirjoitettu hän-muodossa, ja *Stiina* on sen päähenkilö. Ylenmääräistä toistoa ei kuitenkaan juuri ole, sillä kertomus sisältää 229 sanetta.

Koko aineiston 10 yleisintä substantiivilekseemiä kattaa runsaat 16 % ja 30 yleisintä vajaa 30 % kaikista aineiston substantiiveista. 150 yleisintä substantiivilekseemiä kattaa 50 % kaikista aineiston substantiiveista.

Substantiivilekseemejä ei kannata vertailla muiden aineistojen lekseemeihin, sillä aihepiirin vaikutus niihin on merkittävän suuri. *unelma* on hyvä esimerkki lekseemistä, jota ei löydy kovin paljon esimerkiksi sanomalehtitekstistä. CSC:n sanomalehtikielen taajuussanastossa se on sijalla 2 334 vajaan 0,005 %:n osuudellaan. Tässä aineistossa se on kuitenkin toiseksi eniten käytetty substantiivilekseemi ja koko sanastossakin se on 20. yleisin lekseemi.

Vaikka koko aineiston 30 yleisintä substantiivilekseemiä (kuvio 12) löytyvät lähes joka luokalta yleisimpien substantiivien joukosta, jonkin verran vaihtelua niiden osuuksissa on. Edes yleisimmät lekseemit *päivä* ja *unelma* eivät ole eniten käytettyjä substantiiveja joka luokalla. 1. luokan aineistossa muita luokkia selvästi suuremmat osuudet saavat lekseemit *kaveri*, *aamu*, *ilta*, *koira*, *raha*, *karkki* ja *ruoka*. Vaikka voisi kuvitella, että 1.-luokkalaisille äiti ja isä olisivat tärkeämpiä kuin isommille lapsille, *äiti*- ja *isä*-lekseemien osuudet ovat suuremmat lähes kaikilla muilla luokilla. Lekseemejä *hotelli* ja *huone* ei löydy 1.-luokkalaisten kirjoitelmista lainkaan. 1.-luokkalaisten aineistoa tarkasteltaessa on kuitenkin aina huomioitava sen pienuus muiden luokkien aineistoihin verrattuna.



Kuvio 12. Koko aineiston 30 yleisimmän substantiivin suhteelliset osuudet.

30 yleisimmän substantiivilekseeminjoukosta löytyvät kaikki perusvuorokaudenajat, *aamu, päivä, ilta* ja *yö*. Tämä kertoo siitä, että unelmien päivää usein jaksotetaan vuorokaudenaikojen avulla. Tarkempaan ajankäyttöön liittyvät *kello* ja *tunti* kasvattavat osuuksiaan erityisesti 4. luokasta lähtien. Muita selkeästi enemmän vanhempien oppilaiden käyttämiä lekseemejä ovat *matka, auto* ja *ystävä*. Nuorempien oppilaiden suosimia taas ovat *koira, koulu, karkki, kissa* ja *maailma*. Yleisesti ottaen eläimiin viittaavat sanat ovat yleisempiä nuoremmilla kuin varttuneemmilla oppilailla. Hevoset alkavat kuitenkin kiinnostaa enemmän ensimmäisen luokan jälkeen, kunnes kiinnostus näyttää hiipuvan taas 6. luokalla.

Eniten kuviosta pistävät esiin 1.–3.-luokkalaisten palkit, mikä tarkoittaa, että alemmilla luokilla käytetään suhteellisesti enemmän samoja lekseemejä kuin ylemmillä luokilla. Yleisimpien lekseemien osuudet ovat siis alemmilla luokilla suurempia kuin ylemmillä luokilla, joilla lekseemien käyttö hajaantuu enemmän.

Tarkastellaan vielä erikseen 2.-, 4.- ja 6.-luokkalaisten tyttöjen ja poikien eniten käyttämiä appellatiiveja (taulukko 6).

Taulukko 6. Tyttöjen ja poikien 30 frekventeintä appellatiivilekseemiä.

| 2. luokka | | 4. luokka | | 6. luokka | |
|--------------------|--------------------|--------------|---------------|-----------------|------------------|
| tytöt | pojat | tytöt | pojat | tytöt | pojat |
| päivä | päivä | päivä | päivä | päivä | päivä |
| äiti | koulu | äiti | unelma | unelma | unelma |
| koira | koti | isä | koti | äiti | koti |
| koti | unelma | koira | kaveri | ystävä | kello |
| isä | äiti | unelma | tunti | aamu | hotelli |
| nimi | kaveri | aamu | äiti | kello | kaveri |
| aamu | karkki | koti | koulu | matka | aamu |
| kissa | ruoka | hevonon | aamu | koti | raha |
| unelma | aamu | nimi | loppu | hotelli | aika |
| hevonon | loppu | huone | ilta | ilta | ilta |
| kaveri | raha | kaveri | aamupala | kaveri | mopo |
| koulu | isä | kello | kello | aika | tunti |
| tyttö | eläin | ilta | palvelija | tunti | auto |
| koiranpentu | maailma | tunti | pele | aurinko | isä |
| aamupala | pele | aika | isä | koira | pele |
| kauppa | synttärilit | auto | karkki | isä | maali |
| lahja | aika | ystävä | klo | ranta | matka |
| loppu | kissa | aamupala | koira | hevonon | äiti |
| sika | maalivahti | poni | serkku | tyttö | euro |
| talo | vesi | lahja | auto | auto | koulu |
| raha | joulu | palvelija | nimi | kaupunki | ravintola |
| syntymäpäivä | auto | sisko | yö | koulu | tietokone |
| tunti | koira | sänky | loma | aamupala | kauppa |
| isi | loma | talli | raha | elokuva | lotto |

| | | | | | |
|---------------|---------------|--------|-----------------|--------------|--------------------|
| lumi | ottelu | matka | euro | vaate | aamupala |
| matka | poika | karkki | hotelli | mökki | joukkue |
| susi | aamupala | koulu | limsa | sänky | loppu |
| sänky | euro | ruoka | matsi | perhe | perhe |
| maa | huone | loppu | kesäloma | raha | kone |
| peikko | isi | vesi | paikka | ruoka | lentokenttä |
| | kettu | | pari | | talo |
| | metsä | | piha | | vanhemmat |
| | perhe | | raketti | | yö |
| | possu | | sisko | | |
| | syntymäpäivä | | tietokone | | |
| | työ | | vuosi | | |
| | vuosi | | | | |

Taulukon listat sisältävät yhteensä 85 lekseemiä, joista vain yhdeksän löytyy jokaisen vertailuryhmän listalta. Nämä lekseemit ovat aakkosjärjestyksessä *aamu*, *aamupala*, *isä*, *kaveri*, *koti*, *koulu*, *päivä*, *unelma* ja *äiti*. Yhteisistä lekseemeistä neljä liittyy kotiin ja ihmissuhteisiin, lekseemit *aamu* ja *aamupala* ovat tyypillisiä kirjoitelmien aluissa ja *unelma* ja *päivä* liittyvät suoraan aineiden otsikkoon. Koulu on tietenkin tärkeä osa lasten arkea.

Lekseemejä, joita esiintyy vain yhden ryhmän listassa, on 42, ja niistä suurin osa, 28 kappaletta, löytyy poikien listoilta. Yhtenä syynä tähän on se, että pojilla 30 frekventeimmän lekseemin joukkoon kuuluu aina yli 30 lekseemiä.

Useamman kuin yhden luokan tyttöjen listalla on viisi sellaista lekseemiä, *hevonen*, *lahja*, *sänky*, *tyttö* ja *ystävä*, joita ei poikien listoilta löydy. Poikien listoilla taas on kuusi tyttöjen listoilta puuttuvaa lekseemiä. Ne ovat *euro*, *loma*, *pele*, *tietokone*, *vuosi* ja *yö*. Myös muita yleisimpiä lekseemejä esiintyy vain tiettyjen ryhmien listoilla. *syntymäpäivä*, *kissa* ja *isi* löytyvät vain 2.-luokkalaisten listoilta. Nuorimmat oppilaat näyttävät sitovan unelmien päivän vielä herkästi johonkin juhlaan, kuten jouluun tai syntymäpäivään. *sisko* ja *palvelija* esiintyvät ainoastaan 4.-luokkalaisten listoilla ja *palvelijaa* ovat käyttäneet vain kaupunkikoulun oppilaat. *ilta* ja *kello* löytyvät vain 4.- ja 6.-luokkalaisten listoilta. 4.-luokkalaisten poikien listalta löytyy *kellon* lisäksi myös lyhenne *klo*. Tämä voisi kertoa siitä, että tarkan ajan merkitys kasvaa lasten varttuessa.

6.-luokkalaisten poikien listalta ei löydy yhtään eläimiin liittyvää sanaa, vaikka esimerkiksi *koira* löytyy kaikkien muiden ryhmien listoilta. Tyttöjen listoilla se on joka luokalla huomattavasti korkeammalla kuin poikien, mutta jo 6. luokan tyttöjen kirjoitelmissa koiran sijoitus on laskenut alempiin luokkiin verrattuna. Eläinten merkitys unelmissa näyttääkin vanhemmiten vähenevän, ja tilalle tulevat tytöillä elokuvat ja vaatteet, pojilla taas urheilu, mopot, tietokoneet ja muut tekniset vempaimet. Tosin tulosten yleistäminen näin pienessä aineistossa on hieman kyseenalaista. Esimerkiksi 2.-luokkalaisten listalla varsin

korkealle sijoittuneet *sika* ja *peikko* esiintyvät ainoastaan yhdessä, joskaan eivät samassa kirjoittelmassa.

Vähintään viisi kertaa käytettyjä appellatiivilekseemejä on aineistossa 623 kappaletta. Näistä 59:ää ovat käyttäneet vain tytöt ja 26:ta vain pojat. Vain tyttöjen käyttämät lekseemit ovat aakkosjärjestyksessä *aitaus, harja, heinä, henkilö, heppa, hoitohevonen, hytti, häntä, häät, judo, julkkis, juttu, kallio, karsina, kasvattaja, kaura, keppihevonen, kerros, kolo, koralli, koru, kummi, käynti, laidun, laukka, laulu, lava, lempihevonen, maasto, meikki, niitty, olo, omakotitalo, pakko, peikko, poikaystävä, poni, puomi, puuha, pyyhe, ratsastus, ratsastusleiri, riimu, saksanpaimenkoira, satama, satula, sieni, silmälasit, suitset, syli, tarha, tarvike, toppi, traileri, uimapuku, vaatekauppa, varvas, viinirypäle ja yöpuku*. Vain poikien käyttämät lekseemit ovat *aloitus, erä, helikopteri, huvila, juhannus, jääkiekko, jääkiekkjoukkue, jääkiekkomaila, kapteeni, kettu, maalivahti, miljonääri, moottoripyörä, mopo, poliisi, ralli, rikollinen, siivooja, sopimus, stadion, tehdas, traktori, tunneli, tuomari, vapaapäivä ja verryttely*. Kuten voi havaita, selvä enemmistö listojen appellatiiveista on hyvin konkreettisia sanoja. Vain tyttöjen käyttämistä lekseemeistä suuri osa liittyy hevosharrastukseen. Vain poikien käyttämistä lekseemeistä monet taas ovat moottoriajoneuvojen nimiä tai liittyvät urheiluun ja erityisesti jääkiekkoon.

Lopuksi tarkastellaan vielä hieman substantiivilekseemejä, joita on käytetty vain kerran. Monet vain kerran käytetyistä substantiiveista ovat propreja, kahta tavua pidempiä sanoja kuten *arvonta*, tai yhdyssanoja, kuten *asuinalue, päiväpeitto, rikkaruoho, rosvontorjuntajärjestelmä*. Myös joitakin omaperäisiä sanoja löytyy, esimerkkinä *puuroklöntyskä* (esimerkki 6) ja *ruikkunakkiuimahalli* (esimerkki 7).

6) Söin inhoten puuroklöntysköitä. (kielikoulu 3T69)

7) Hän oli uskomaton mutta totta siis eli ruikkunakki joka meni ruikkunakkiuimahalli (kaupunki 2P160)

Koska substantiivit ovat sanaluokkana erittäin produktiivinen, vain kerran käytettyjä lekseemejä on substantiivilekseemeistä noin 58 %. Runsaat 75 % oppilaista on käyttänyt ainakin yhtä vain kerran esiintyvää substantiivilekseemiä. Yhden kerran käytettyjä lekseemejä eli HL-sanoja käsitellään enemmän luvussa 3.3.

3.2.2 Verbit

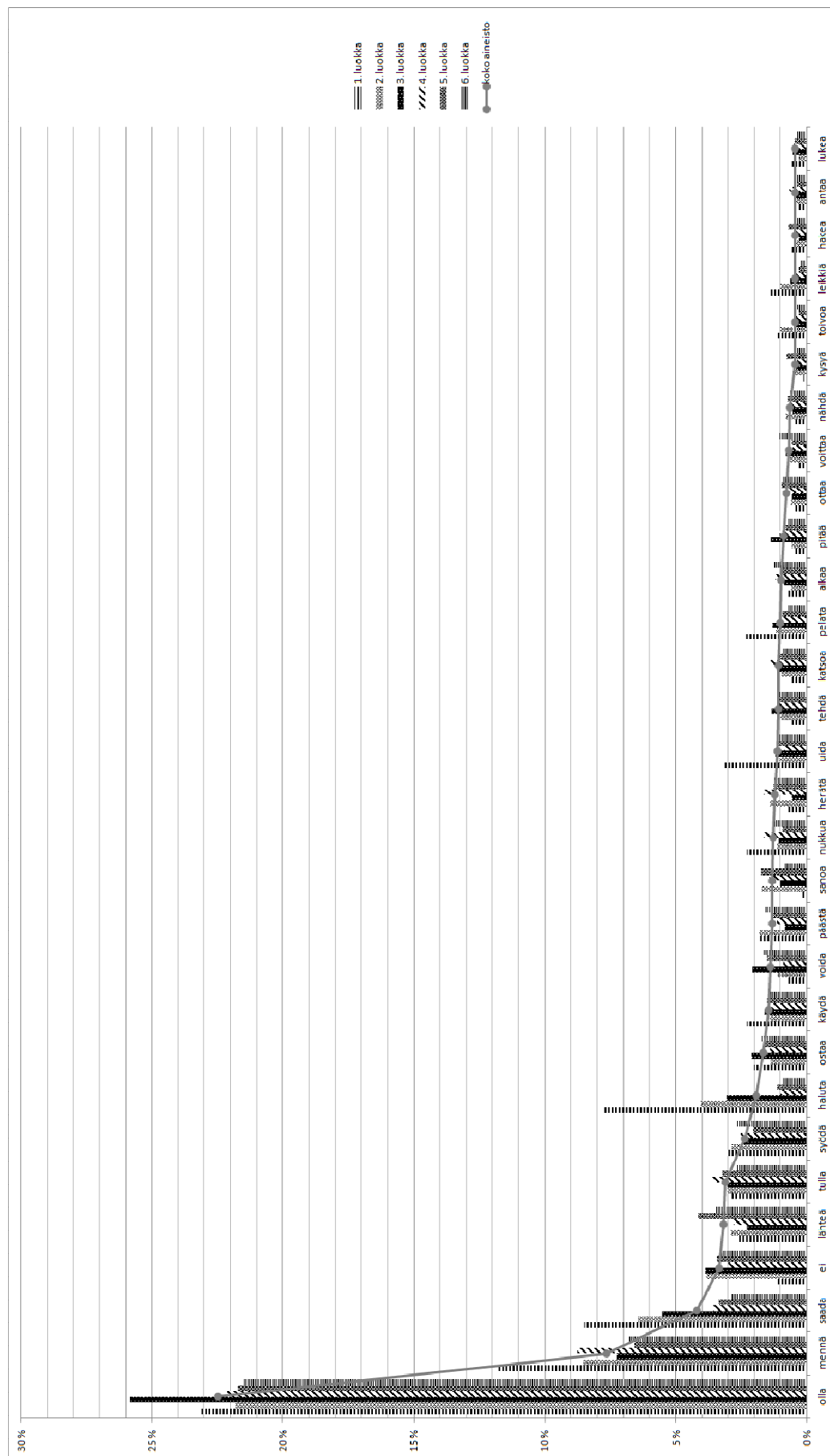
Aiemmissa Koulukorpusta aineistona hyödyntävissä tutkimuksissa on jo selvitetty alakouluikäisten verbien käyttöä muun muassa semanttisten verbiluokitusten avulla (Pajunen 2012: 21). Verbit on jaettu tekoja, liikettä, tapahtumia ja tiloja merkitseviin verbeihin eli A-verbeihin, fysiologisia tiloja ja prosesseja sekä verbaalitekoja ilmaiseviin A/B-verbeihin, havaintoja, emootiota, propositionaalisia asenteita ja puheakteja kielentäviin B-verbeihin ja kieliopillista merkitystä kantaviin verbeihin. Tulosten perusteella A-verbien osuus pienenee 1. luokalta 6. luokalle, mikä kertoo semanttisesti kompleksisempia ja abstraktimpia merkityksiä koodaavien verbilekseemien lisääntymisestä teksteissä. Kieliopillista merkitystä kantavien verbien lisääntyminen kertoo muun muassa perfektin- ja pluskvamperfektimuotojen lisääntymisestä. (mts.) Tässä tutkimuksessa semanttisia luokitteluja ei siis enää käytetä hyväksi verbienkäyttöä kartoitettaessa.

Aineiston kymmenen frekvenssiltään yleisintä verbiä ovat *olla, mennä, saada, ei, lähteä, tulla, syödä, haluta, ostaa* ja *käydä*. Nämä kymmenen suurifrekventtisintä verbilekseemiä kattavat yli 50 % kaikista kirjoitelmissa käytetyistä verbeistä. 30 eniten käytetyn verbilekseemin (kuvio 13) osuus kaikista verbeistä on lähes 70 %.

Verbit, joiden osuus on selkeästi suurempi alemmilla kuin ylemmillä vuosiluokilla, ovat *mennä, saada, haluta, pelata, toivoa* ja *leikkiä*. Osittain tämä johtuu siitä, että iän myötä sanavarasto karttuu ja yleisimmät verbit eivät enää kata yhtä suurta osuutta aineistosta kuin alemmilla luokilla. Yleisimmät 30 verbiä voivat kertoa jotakin myös kirjoittajien ajatusmaailmasta. On selvää, että ainakin leikkiminen vähenee lapsen varttuessa. Ylemmillä vuosiluokilla selvästi suuremman osuuden anastavat verbit *lähteä, alkaa, ottaa* ja *voittaa*. Mielenkiintoinen on kontrasti saamisen ja ottamisen välillä. Nuoremmat lapset tyytyvät vielä enemmän siihen, mitä saavat, kun vanhemmat lapset jo useammin ottavat mitä haluavat.

Verbien vertailu muihin aineistoihin on helpompaa kuin adjektiivien ja erityisesti substantiivien, sillä aihepiiri ei vaikuta yhtä paljon varsinkaan kirjoitelmissa eniten käytettyihin verbeihin, jotka kuuluvat perussanastoon. Koulukorpuksen 30 yleisimmän verbin joukossa on yhdeksän sellaista verbiä, joita ei löydy Opettajan sanastokirjan (1970) 30 yleisimmän verbin joukosta. Ne ovat yleisyysjärjestyksessä *haluta, nukkua, herätä, uida, pelata, voittaa, kysyä, toivoa* ja *hakea*. Kyseisistä verbeistä nukkua ja herätä liittyvät aineiden

tavalliseen kertomusrakenteeseen, sillä aineet alkavat usein siitä, kuinka aamulla herätään ja loppuvat siihen, kun illalla mennään nukkumaan. Haluaminen ja toivominen ovat myös unelmien päivälle tyypillisiä. Uiminen, pelaaminen ja pelaamiseen liittyvä voittaminen kuuluvat myös usein unelmien päivään. Vaikka Opettajan sanastokirjan sanasto onkin kerätty aineista, jotka on kirjoitettu monenlaisista otsikoista, yleisimpien verbien joukossa on mielestäni yllättävänkin paljon samoja verbejä, mikä osaltaan kertoo siitä, että jonkinlaisella perusverbivarastolla voi kirjoittaa hyvin monesta aiheesta. Eniten 40 vuoden takaisesta listasta hyppää esiin verbi *panna* (sija 19), joka ei Koulukorpuksessa esiinny edes sadan yleisimmän verbilekseemin joukossa. Eräs mielenkiintoinen muutos 40 vuoden takaisiin koululaisiin onkin nähtävissä verbien *panna*, *laittaa* ja *pistää* käytössä. Opettajan sanastokirjassa verbi *panna* on sadan yleisimmän lekseemin joukossa, sijalla 78, ja *laittaa* on 97. yleisin lekseemi. Koulukorpus-aineistossa näistä kolmesta yleisin verbi on *laittaa* (frekvenssi 0,1 % saneista), sitten on *pistää*, ja *panna*-verbiä on käytetty vain neljä kertaa. Näiden verbien kohdalla voi todeta, että ennen neutraali *panna* ei ole enää neutraali, vaan sen tilalla käytetään ensisijaisesti verbiä *laittaa*, joka yhdistetään ainakin hyvässä yleiskielessä ennen kaikkea ruoanlaittoon.



Kuvio 13. Koko aineiston 30 yleisimmän verbin suhteelliset osuudet.

Mitä verbejä tytöt käyttävät mutta pojat eivät ja päinvastoin? Koko aineistossa vähintään kolme kertaa esiintyvien verbien (442 lekseemiä, 96 % saneista) joukossa on 74 sellaista lekseemiä, jota käyttävät vain joko tytöt tai pojat. Vain tyttöjen aineissa esiintyvät 63 verbiä ovat *esiintyä, haista, halata, harjata, heilua, heiluttaa, hengittää, herkutella, hoputtaa, hymyillä, hyvästellä, hämmästyä, iloita, istahtaa, jatkua, johdattaa, kantaa, karata, katsahtaa, keilata, kiinnittää, kiirehtiä, kiiruhtaa, kiitellä, kiljua, kuiskata, kuvitella, kyllästyä, laahustaa, laukata, leijua, lennellä, loistaa, lomailla, lueskella, malttaa, meikata, moikata, muistella, nuolla, nyökätä, näytellä, paketoita, panna, pissata, polskutella, pussailla, päätyä, rapsuttaa, riidellä, ruskettua, sitoa, soutaa, suitsia, suunnitella, syttyä, taluttaa, tanssia, tipahtaa, tuijottaa, uiskennella, varustaa, välittää*. Vain poikien aineissa esiintyvät 11 verbiä ovat *erota, korjata, laukoa, lihoa, lyödä, räjäyttää, skeitata, sortua, taistella, torjua ja urheilla*.

Tyttöjen käyttämiin verbeihin kuuluu niin sanottuja hellittelyverbejä, kuten *halata, pussailla ja rapsuttaa*, ulkonäköön liittyviä kuten *meikata ja ruskettua* sekä emotionaalisia verbejä kuten *iloita, kyllästyä ja välittää*. Useat vain tyttöjen käyttämät verbit ovat johdettuja. Vain poikien käyttämät verbit liittyvät usein urheiluun, kuten *laukoa, lyödä, skeitata, torjua ja urheilla* tai muuten miehiseen maailmaan, kuten *räjäyttää ja taistella*, joita toki urheilukielessäkin käytetään. Ehkä hieman yllättäenkin yleisimpien verbien listalla sijalle 91 yltävä puhekielinen verbi *shoppailla* esiintyy sekä tyttöjen että poikien kirjoitelmissa.

Yhden kerran käytetyistä verbeistä 371 löytyy tytöiltä ja 175 pojilta. Yhteensä koko aineistossa on 528 sellaista verbiä, jota käyttävät vain tytöt ja 224 sellaista verbiä, jota käyttävät vain pojat. Saneiden määrään suhteutettuna ”tyttömäisiä” verbejä on tyttöjen käyttämistä verbeistä 3,6 % ja ”poikamaisia” verbejä poikien käyttämistä verbeistä 2,4 %. Näiden lukujen perusteella tyttöjen verbien käyttö vaikuttaa huomattavasti poikia monipuolisemmalta.

Seuraavissa taulukoissa (taulukot 7 ja 8) esitetään joitakin perustietoja 2.-, 4.- ja 6.-luokkalaisten verbien käytöstä.

Taulukko 7. Tietoja 2., 4. ja 6. luokan verbienkäytöstä.

| | 2. luokka | 4. luokka | 6. luokka |
|--------------------------|-----------|-----------|-----------|
| verbisaneita yht. | 3049 | 5521 | 5000 |
| verbilekseemejä yht, | 303 | 548 | 550 |
| HL-verbejä yht. | 128 | 278 | 266 |
| HL-verbit % lekseemeistä | 42 | 51 | 48 |

Taulukko 8. 2.-, 4.- ja 6.-luokkalaisten tyttöjen ja poikien verbienkäyttö.

| | 2. luokka | | 4. luokka | | 6. luokka | |
|---|-----------|-------|------------|------------|-----------|-------|
| | tytöt | pojat | tytöt | pojat | tytöt | pojat |
| osuus luokan verbisaneista (%) | 61 | 39 | 64 | 36 | 61 | 39 |
| verbisaneita aineessa keskimäärin (kpl) | 16,3 | 10,8 | 30,2 | 20,2 | 39,3 | 23,6 |
| verbien osuus saneista (%) | 29 | 30 | 30 (29)* | 28 (28)* | 29 | 28 |
| verbilekseemin keskimääräinen toisto | 7,7 | 6,8 | 7,8 (6,6)* | 7,3 (6,5)* | 6,9 | 6,7 |
| osuus luokan HL-verbeistä (%) | 64 | 36 | 70 | 30 | 71 | 29 |
| käyttäneet % kaikista luokan verbilekseemeistä | 80 | 57 | 83 | 49 | 81 | 53 |
| % oppilaista käyttänyt verbiä, jota kukaan muu ei käytä | 36 | | 61 | | 62 | |

*Luku ei sisällä kielikylpykirjoitelmia.

Vertaillaan vielä hieman 2., 4. ja 6. -luokkien verbien käyttöä. Verbisaneiden määrä yhden oppilaan aineessa keskimäärin näyttää tytöillä lisääntyvän huomattavasti toiselta luokalta kuudennelle, kun taas pojilla verbien määrä kaksinkertaistuu toiselta neljännelle, mutta neljänneltä kuudennelle määrä lisääntyy enää kolmella verbillä (taulukko 8). Verbien osuudessa saneista ei juurikaan tapahdu muutosta luokka-asteiden välillä. Kielikylpykoulun tyttöillä verbien osuus on suurempi kuin muiden koulujen tytöillä, mistä ainakin osittain seuraa myös se, että kielikylpykoululaisten mukana olo aineistossa lisää verbien toistoa. Toisaalta toisto lisääntyy myös siksi, että aineiston koko eli saneiden määrä kasvaa, kun kielikylpykirjoitelmat ovat mukana aineistossa. Koko luokan verbisaneista suurin osa esiintyy kaikilla luokka-asteilla tyttöjen kirjoitelmissa, vaikka kuudennella luokalla tyttökirjoittajia on vähemmän kuin poikakirjoittajia.

Koska oletuksena on, että jo pieni määrä oppilaita käyttää suurinta osaa perusverbeistä, tarkastellaan seuraavaksi sitä, kuinka suurta osaa eniten verbejä aineisiinsa sisällyttäneet oppilaat ovat koko luokan verbeistä käyttäneet.

2. luokalla kahdestatoista eniten verbejä käyttäneestä oppilaasta (29–38 kappaletta) yhdeksän oli tyttöjä ja kolme poikia. Yhteensä he käyttivät 102 eri verbilekseemiä, mikä on 34 % 2.-luokkalaisten käyttämistä verbilekseemeistä. Heidän käyttämänsä verbit kattavat 13 % kaikista 2.-luokkalaisten käyttämistä verbisaneista.

Enimmillään verbejä esiintyi 38 kappaletta yhden tytön ja yhden pojan kirjoitelmassa. Tyttö käytti 21 eri verbilekseemiä ja poika 16:tta. Tytön kirjoitelmassa verbit

kattavat 33 % saneista ja pojan kirjoitelmassa 39 %. Tytön kirjoitelman tasoksi on arvioitu kertomus ja pojan ei skeemaa.

4. luokalla kymmenestä eniten verbejä käyttäneestä oppilaasta (56–83 kappaletta) yhdeksän oli tyttöjä ja yksi poika. Tytöistä seitsemän oli kielikylpykoululaisia. Yhteensä nämä kymmenen oppilasta käyttivät 150 eri verbilekseemiä, mikä kattaa 27 % 4.-luokkalaisten aineissa esiintyneistä verbilekseemeistä, ja heidän käyttämänsä verbit kattavat 11 % kaikista 4.-luokkalaisten käyttämistä verbisaneista (kymmenen oppilasta kattaa vajaat 5 % kaikista 4.-luokan kirjoittajista).

Enimmillään verbejä esiintyi 83 kappaletta yhden kielikylpykoulutytön kirjoitelmassa. Tyttö käytti 41 eri verbilekseemiä. Toiseksi eniten verbejä (68 kpl) käytti tavallisesta koulusta oleva tyttö. Hän käytti 40 eri verbilekseemiä. Eniten verbejä käyttänyt poika (58 kpl) käytti 24 eri verbilekseemiä. Tavallisista koululaisista tytön kirjoitelman tasoksi on arvioitu kertomus ja pojan skripti. Tytön kirjoitelmassa verbit kattavat 32 % saneista ja pojan kirjoitelmassa 29 %. Kielikylpykoulutytön kirjoitelmassa verbit kattavat saneista 37 %.

6. luokalla yhdestätoista eniten verbejä (59–105 kappaletta) käyttäneestä oppilaasta kymmenen oli tyttöjä ja yksi poika. Yhteensä he käyttivät 234 eri verbilekseemiä, jotka kattavat kaikista 6.-luokkalaisten käyttämistä verbilekseemeistä 43 %, ja heidän käyttämänsä verbit kattavat 16 % kaikista 6.-luokkalaisten käyttämistä verbisaneista.

Enimmillään verbejä esiintyi 105 kappaletta yhden tytön kirjoitelmassa. Eniten verbejä käyttäneellä pojalla niitä oli 59 kappaletta. Tyttö käytti 51 eri verbilekseemiä ja poika 41:tä, mikä tarkoittaa, että keskimäärin työllä esiintyi verbilekseemin toistoa huomattavasti enemmän kuin pojalla. Tytön kirjoitelmassa verbit kattavat 29 % saneista ja pojan kirjoitelmassa 42 %. Molempien kirjoitelmien tasoksi on arvioitu skripti.

Vertailemalla 2., 4. ja 6. luokan oppilaita voi havaita muutamia selkeitä muutoksia: Verbejä esiintyy joka luokalla enemmän tyttöjen kuin poikien kirjoitelmissa, ja ero vain kasvaa, kun siirrytään ylemmille vuosiluokille. Osittain verbien määrä kertoo kirjoitelmien pituudesta: pitäisihän jokaisen lauseen sisältää yksi finiittiverbi, joten mitä enemmän verbejä esiintyy, sitä pidemmistä kirjoitelmista on kyse. Näin yksioikoinen asia ei tietysti ole. Verbien määrän lisääntyminen voi merkitä myös sitä, että tytöt käyttävät monimutkaisempia tempuksia ja rakenteita kuin pojat. Ainakin partisiippimuotojen osuus on tyttöjen kirjoitelmissa

huomattavasti suurempi kuin poikien (taulukko 9) sekä toisella että kuudennella luokalla. 4. luokka poikkeaa 2. ja 6. luokan trendistä.

Taulukko 9. Verbin infinitiivi- ja partisiippimuotojen osuus verbisaneista.

| | 2. luokka | | 4. luokka | | 6. luokka | |
|---------------------|-----------|-------|-----------|-------|-----------|-------|
| | tytöt | pojat | tytöt | pojat | tytöt | pojat |
| infinitiivimuodot % | 14 | 19 | 11 | 10 | 16 | 19 |
| partisiippimuodot % | 5 | 2 | 4 | 4 | 8 | 5 |

Ylemmillä luokilla useammat oppilaat käyttävät sellaista verbilekseemiä, jota kukaan muu ei käytä. Tämä kertoo ainakin siitä, että oppilaiden käyttämä sanasto on eriytynyt. Niilläkin luokilla, joilla tyttöjä on vähemmän kuin poikia, tytöt käyttävät verbilekseemeistä aina huomattavasti suurempaa osaa kuin pojat. Kuten aiemminkin on todettu, saneiden määrä on selvästi yhteydessä toiston määrään erityisesti alemmilla luokilla; mitä enemmän saneita, sitä enemmän toistoa. 6.-luokalla saneiden ja toiston yhteys ei enää näytä yhtä selvältä. Vaikka verbisaneita onkin tyttöjen kirjoitelmissa huomattavasti enemmän kuin poikien, verbilekseemien keskimääräistä toistoa tyttöjen teksteissä esiintyy enää hyvin vähän enemmän kuin poikien teksteissä. Edellä esitettyjen karkeiden muuttujien perusteella näyttää siltä, että tyttöjen käyttämä verbisanasto on huomattavasti poikia monipuolisempaa, ja tyttöjen ja poikien välinen ero vain kasvaa sitä mukaa mitä ylemmälle luokalle siirrytään.

Vain kerran käytettyjä lekseemejä on verbilekseemeistä noin 47 %. Vain 28 % oppilaista on käyttänyt ainakin yhtä vain kerran esiintyvää verbilekseemiä, kun osuus substantiivien kohdalla oli yli 75 %. Tämä kertoo ensinnäkin siitä, että vaikka aineet yhteisestä otsikosta huolimatta saattavat kuvata unelmien päivää ja sen tapahtumia hyvin eri tavoin, verbisanastoon tämä vaihtelu ei vaikuta yhtä suuresti kuin substantiivisanastoon. Omaperäisiä tai ekspressiiviä verbejä ei myöskään keksitä tai ainakaan käytetä yhtä herkästi kuin omaperäisiä substantiiveja. Yhden kerran käytettyjä lekseemejä eli HL-sanoja käsitellään vielä luvussa 3.4.1.

Seuraavaksi tarkastellaan 2.-, 4.- ja 6.-luokkalaisten tyttöjen ja poikien 30:tä eniten käyttämää verbilekseemiä (taulukko 10). Vaikka listaa käsitelläänkin itsenäisesti muusta aineistosta irrallaan, on syytä muistaa, että nekin ryhmät, joiden listoilta tiettyjä verbejä ei löydy, ovat saattaneet niitä käyttää, vaikka kyseiset lekseemit eivät olekaan mahtuneet 30 eniten käytetyn joukkoon.

Taulukko 10. Tyttöjen ja poikien 30 frekventeintä verbilekseemiä.

| 2. luokka | | 4. luokka | | 6. luokka | |
|---------------|-------------------|------------------|--------------|----------------|----------------|
| tytöt | pojat | tytöt | pojat | tytöt | pojat |
| olla | olla | olla | olla | olla | olla |
| mennä | saada | mennä | mennä | mennä | mennä |
| saada | mennä | tulla | saada | ei | saada |
| ei | haluta | ei | tulla | lähteä | lähteä |
| haluta | ei | saada | lähteä | saada | tulla |
| lähteä | syödä | lähteä | syödä | syödä | syödä |
| tulla | pelata | syödä | ei | tulla | ostaa |
| syödä | tulla | sanoa | ostaa | voida | ei |
| sanoa | päästä | herätä | pelata | alkaa | voittaa |
| herätä | lähteä | nukkua | nukkua | nukkua | päästä |
| päästä | ostaa | katsoa | herätä | päästä | käydä |
| käydä | toivoa | käydä | tehdä | herätä | voida |
| uida | nukkua | pitää | alkaa | käydä | katsoa |
| ostaa | käydä | ostaa | katsoa | ostaa | nukkua |
| leikkiä | tehdä | voida | päästä | uida | pelata |
| voida | voida | alkaa | käydä | haluta | tehdä |
| nukkua | sanoa | haluta | haluta | ottaa | herätä |
| katsoa | katsoa | päästä | voittaa | tehdä | alkaa |
| nähdä | herätä | ottaa | uida | sanoa | ottaa |
| tehdä | osata | uida | antaa | nähdä | sanoa |
| kysyä | leikkiä | tehdä | sanoa | katsoa | haluta |
| toivoa | lentää | nähdä | tuoda | pitää | uida |
| voittaa | ajaa | laittaa | ajaa | odottaa | pitää |
| ottaa | alkaa | antaa | pitää | palata | ajaa |
| pitää | nähdä | lukea | nähdä | saapua | soittaa |
| asua | laskea | odottaa | odottaa | kysyä | huomata |
| alkaa | lasketella | kysyä | voida | laittaa | nähdä |
| löytää | uida | kävellä | hakea | päättää | viedä |
| antaa | muuttaa | ruveta | juoda | viedä | antaa |
| hakea | ottaa | hakea | ottaa | lukea | avata |
| juosta | tarvita | ratsastaa | | soittaa | laittaa |
| | voittaa | | | | lentää |
| | | | | | odottaa |
| | | | | | palata |
| | | | | | tarvita |

Yhteen laskettuna kaikkien ryhmien 30 frekventeimmän verbin listalla on 54 eri verbilekseemiä. Listalla on 21 sellaista lekseemiä, jotka ovat kaikkien ryhmien 30 yleisimmän verbilekseemin joukossa. Nämä lekseemit ovat aakkosjärjestyksessä *alkaa*, *ei*, *haluta*, *herätä*, *katsoa*, *käydä*, *lähteä*, *mennä*, *nukkua*, *nähdä*, *olla*, *ostaa*, *ottaa*, *päästä*, *saada*, *sanoa*, *syödä*, *tehdä*, *tulla*, *uida* ja *voida*. Näistä verbeistä kaikki muut paitsi *herätä*, *katsoa*, *nukkua*, *ostaa*, *syödä* ja *uida* löytyvät myös CSC:n Suomen sanomalehtikielen taajuussanaston 30 yleisimmän verbilekseemin joukosta. Verbeistä *herätä* ja *nukkua* ovatkin ehkä tyypillisempiä tarinamuotoiselle kirjoitelmalle kuin sanomalehtitekstille, *ostaminen* ja *uiminen* liittyvät erityisesti Unelmien päivä -otsikkoon. Kaikilla luokilla kieltoverbi *ei* on tyttöjen listalla

korkeammalla sijalla kuin poikien. Sama tulos näkyy myös Saarelan (1991: 35) tutkimuksissa. Poikien kirjoitelmissa kieltoverbin sijoitus frekventeimpien verbien listalla laskee ylemmillä vuosiluokilla, kun se tyttöillä sen sijaan nousee. Tämä voi johtua siitä, että tytöt käyttävät poikia enemmän monimutkaisempia verbirakenteita ja että kirjoitelman aihetta ei lähestytä kiellon kautta. Monimutkaisten rakenteiden selitys on kuitenkin ongelmallinen siksi, että nuoremmilla pojilla *ei* on frekventimpi kuin vanhemmilla. Sitä voi kuitenkin selittää nuorempien poikien huomattavasti kapeammalla verbivalikoimalla, mistä syystä *ei* harvemmin käytettynä sijoittuu korkealle yleisimpien verbien listalla.

Vain yhden ryhmän listalta löytyviä leksemejä on 16. Ne ovat *asua*, *löytää* ja *juosta* (2. lk tytöt), *osata*, *laskea*, *lasketella* ja *muuttaa* (2. lk pojat), *kävellä*, *ruveta* ja *ratsastaa* (4. lk tytöt), *tuoda* ja *juoda* (4. lk pojat), *saapua* ja *päättää* (6. lk tytöt) sekä *huomata* ja *avata* (6. lk pojat).

Mielenkiintoisimpia ovat kenties verbit, jotka ovat selkeästi keskittyneet vain useamman luokan tytöille tai pojille tai tietylle luokalle. Vain tyttöjen listoilta löytyviä verbejä ovat *kysyä* ja *lukea* ja poikien listoilta löytyviä *ajaa*, *lentää*, *pelata* ja *tarvita*. Nämä verbit tuntuvat kertovankin jotakin tyttöjen ja poikien välisen maailman eroista: tytöt kyselevät ja lukevat, poikien ajanvieton tavat ovat vauhdikkaampia. Kulkemisen tavoista tyttöillä kertoo jotakin jo sekin, että *kävellä* löytyy 4.-luokkalaisten tyttöjen listalta, mutta *ajaa* ja *lentää* vain kaikkien poikien listoilta. 2.-luokkalaisten listoilta löytyvät *leikkiä* ja *toivoo* ja 6.-luokkalaisten *palata*, *soittaa* ja *viedä*. Leikkiminen on tietysti ominaisempaa pienemmille kuin isommille oppilaille, ja toivominen saattaa kertoa siitä, että ollaan toisaalta vielä kiinni satujen maailmassa, jossa tehdään toivomuksia, toisaalta siitä, että ollaan isompia oppilaita enemmän vanhempien holhouksen alaisina, eikä edes tarinassa voida toimia yhtä itsenäisesti kuin vaikkapa 6.-luokkalaiset. 6.-luokkalaiset käyttivät *soittaa*-verbiä enimmäkseen merkityksessä *soittaa puhelimella*, mutta tytöt käyttivät sitä myös muutaman kerran merkityksessä *soittaa jotakin instrumenttia*. Vaikuttaisi siis siltä, että tytöille musiikin merkitys on ehkä suurempi jo kuudennella luokalla kuin pojille, mutta muutaman löydöksen perusteella kovin varmoja päätelmiä ei voi tehdä.

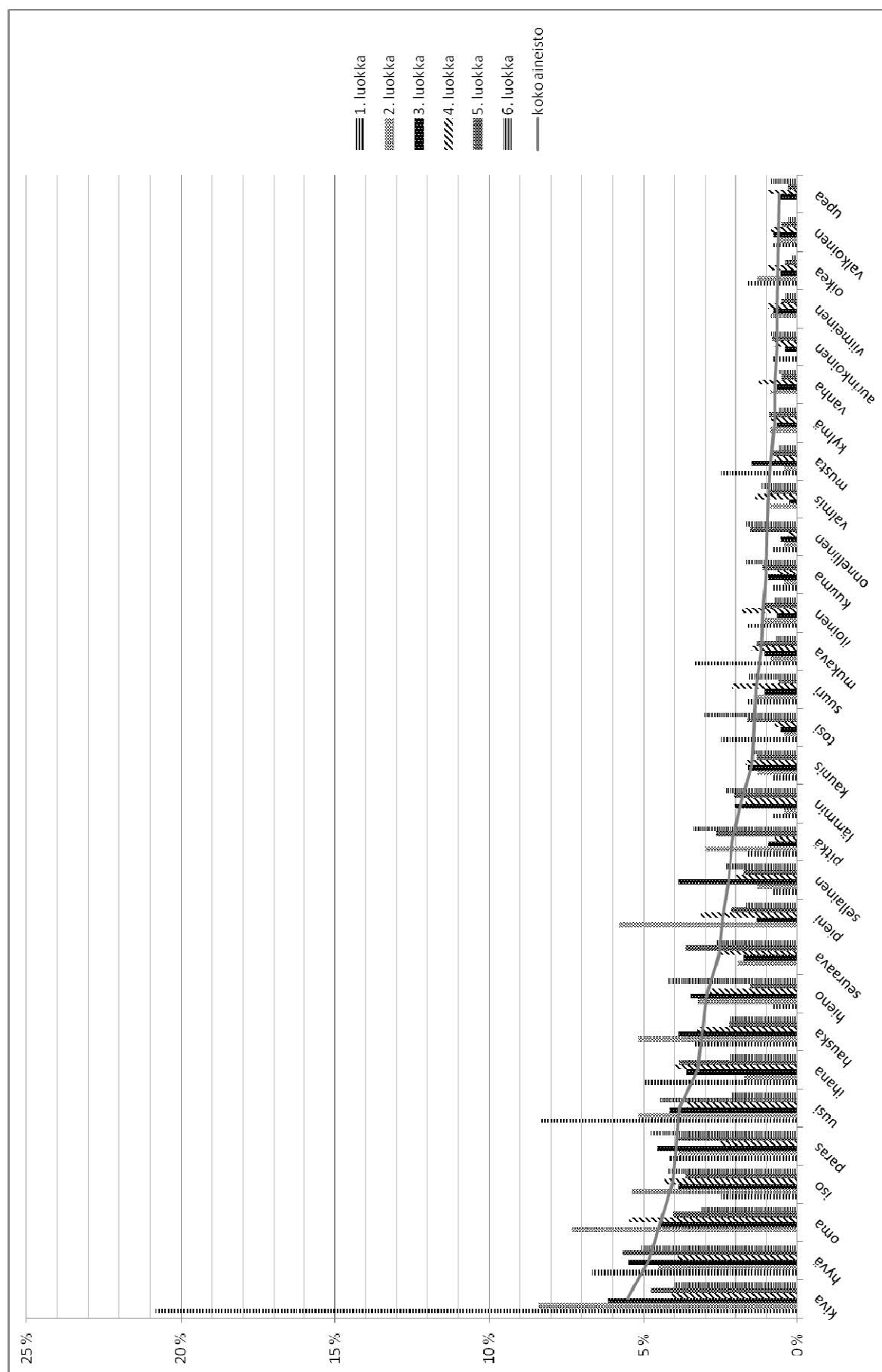
4.- ja 6.-luokkalaisten tyttöjen listoilta ei löydy *voittaa*-verbiä. Voittaminen liittyy tietysti usein pelaamiseen, ja *pelata*-verbiä ei löydy tyttöjen listoilta. Yllättävää ehkä on, että *voittaa* on 2.-luokkalaisten tyttöjen listalla korkeammalla kuin 2.-luokkalaisten poikien.

3.2.3 Adjektiivit

Adjektiivit kuuluvat nominien luokkaan ja ne ilmaisevat esimerkiksi olioiden tai asiantilojen ominaisuuksia (VISK § 603). Alakouluikäisten ominaisuudenilmausten käyttöä on Koulukorpusta aineistonaan käyttäen aiemmin tutkinut Kuisma (2011). Peltonen (2011) on tutkinut lasten kuvakirjoissa esiintyviä adjektiiveja. Tässä tutkimuksessa adjektiiveja tarkastellaan lähinnä lekseemitasolla, ja niiden käyttöä verrataan hieman Peltosen ja Opettajan sanastokirjan tuloksiin. Aluksi luodaan yleiskatsausta Koulukorpus-aineiston adjektiiveihin, sitten vertaillaan adjektiivien käyttöä 2.-, 4.- ja 6.-luokkalaisten teksteissä ja lopuksi tarkastellaan vielä erikseen tyttöjen ja poikien adjektiivien käyttöä

Koko Koulukorpus-aineiston kymmenen yleisintä adjektiivilekseemiä kattaa adjektiiveista lähes 40 % ja 30 yleisintä hieman yli 60 %. Aineiston 30 yleisintä adjektiivia sekä niiden frekvenssit luokittain on esitetty kuviossa 14 ja lista koko aineiston sadasta yleisimmästä adjektiiveista löytyy liitteestä 5.

Opettajan sanastokirjassa (1970: 43) todetaan, että adjektiivivalikoiman kasvaminen on yhteydessä kielen kehittymiseen. 2.-luokkalaisten ovat käyttäneet kirjoitelmissaan yhteensä 130 adjektiivilekseemiä, 4.-luokkalaisten 233:a ja 6.-luokkalaisten 247:ä. Kun adjektiivilekseemien määrä suhteutetaan luokkien kirjoitelmien sanemäärään, saadaan 2.-luokkalaisten suhdeluvuksi 0,012, 4.-luokkalaisten suhdeluvuksi 0,012 ja 6.-luokkalaisten 0,014. Yhtä oppilasta kohti adjektiivilekseemien määrä kasvaa selvästi: 2. luokalla lekseimejä on yhtä oppilasta kohti 0,6, 4. luokalla 1,1 ja 6. luokalla 1,5. Tämän perusteella adjektiivilekseemien määrä kasvaa selvästi ylemmillä vuosiluokilla, joskin kasvu johtuu oppilaiden vartumisesta eikä se ole suorassa yhteydessä kirjoitelmien saneiden määrään.



Kuvio 14. Koko aineiston 30 yleisimmän adjektiivin suhteelliset osuudet.

Koko aineiston 30 yleisimmän adjektiivin joukossa (kuvio 14) huomiota kiinnittää erityisesti se, kuinka suuren osan jo viisi 1.-luokkalaisten eniten käyttämää adjektiivia kattaa kaikista 1.-luokkalaisten käyttämistä adjektiivisaneista. 1.-luokkalaisten eniten käyttämä adjektiivi *kiva* kattaa yli viidenneksen kaikista heidän käyttämistään adjektiiveista. Seuraavaksi yleisimmät adjektiivit ovat *uusi*, *hyvä*, *ihana* ja *paras*. Yhteensä nämä kaikki viisi kattavat 1.-luokkalaisten käyttämistä adjektiivisaneista noin 45 %. Vertailun vuoksi todettakoon, että 3.-luokkalaisten viisi yleisintä adjektiivilekseemiä kattaa runsaat 26 % ja 5.-luokkalaisten enää 23 % kaikista kirjoitelmien adjektiivisaneista. 3.-luokkalaisten viisi yleisintä adjektiivilekseemiä ovat samat kuin 1.-luokkalaistenkin, mutta 5.-luokkalaisten viiden yleisimmän joukossa on nyt adjektiivi *oma*, joka on syrjäyttänyt lekseemin *ihana*. *kiva* ei ole enää 5.-luokkalaisilla eniten käytetty adjektiivilekseemi, vaan frekventein adjektiivi heidän kirjoitelmissaan on *hyvä*.

Adjektiivin *kiva* käytössä näyttäisi tapahtuneen ekspansio viime vuosikymmenten aikana. Opettajan sanastokirjassa (1970) *kiva* on yleisimpien lekseemien listassa sijalla 432, kun taas Koulukorpuksessa se sijoittuu yleisimpien lekseemien listalla sijalle 47.

Vaikka 1.-luokkalaisten aineisto on pieni, on silti hieman yllättävää, että heidän kirjoitelmistaan ei löydy lainkaan useita sellaisia adjektiiveja, jotka ovat koko Koulukorpus-aineistossa 30 yleisimmän joukossa. Nämä adjektiivit ovat *oma*, *seuraava*, *pieni*, *valmis*, *kylmä*, *vanha*, *viimeinen* ja *upea*. Adjektiivi *aurinkoinen* puuttuu täysin 2.-luokkalaisten aineistosta.

Verrattaessa Koulukorpus-aineiston adjektiiveja Peltosen (2011) kuvakirja-aineiston yleisimpiä adjektiiveja, voi havaita joitakin eroavaisuuksia. Koulukorpus-aineiston 53 yleisimmän adjektiivin joukossa on 29 adjektiivia, joita ei löydy Peltosen aineiston yleisimpien adjektiivien joukosta. Tällaisia adjektiiveja ovat esimerkiksi *aurinkoinen*, *onnellinen*, *kauhea* ja *söpö* sekä värit *musta* ja *valkoinen*. Sen sijaan, jos verrataan Koulukorpuksen 20 yleisintä adjektiivia Peltosen 53 yleisimmän adjektiivin joukkoon, niistä vain kaksi, *paras* ja *tosi*, eivät löydy Peltosen listalta. Aineistojen yhteiset adjektiivit ovat enimmäkseen lyhyitä ja niiden voi katsoa kuuluvan perussanastoon.

Koulukorpus-aineiston yleisimmät adjektiivit ovat sävyiltään enimmäkseen varsin positiivisia tai neutraaleja. Yleisimpien adjektiivien (liite 5) joukossa on vain kuusi adjektiivia, *outo*, *kauhea*, *huono*, *tylsä*, *vaikea* ja *paha*, jotka voisi kontekstia sen tarkemmin tutkimatta

luokitella varsin negatiivissävytteisiksi. Epäilemättä aineiden aihe vaikuttaa adjektiivien käyttöön melko lailla, sillä useimmissa kirjoitelmissa unelmien päivään kuuluu lähinnä mukavia asioita.

Seuraavaksi vertaillaan 2., 4. ja 6. luokkien adjektiivien käyttöä (taulukko 11).

Taulukko 11. 2., 4. ja 6.-luokan adjektiivien käytön vertailua.

| | 2. luokka | | 4. luokka | | 6. luokka | |
|--|-----------|-------|-----------|-------|-----------|-------|
| | tytöt | pojat | tytöt | pojat | tytöt | pojat |
| osuus luokan adjektiivisaneista (%) | 64 | 36 | 68 | 32 | 64 | 36 |
| adjektiivisaneita aineessa keskimäärin (kpl) | 2,6 | 1,5 | 5,5 | 3,1 | 8,4 | 4,5 |
| adjektiivien osuus saneista (%) | 4,6 | 4,2 | 5,4 | 4,5 | 6,1 | 5,4 |
| adjektiivileksemin keskimääräinen toisto | 3,0 | 2,6 | 3,5 | 2,7 | 3,4 | 3,0 |
| osuus luokan HL-adjektiiveista (%) | 64 | 36 | 68 | 32 | 66 | 34 |
| käyttäneet % kaikista luokan adjektiivileksemeistä | 74 | 49 | 83 | 49 | 79 | 49 |
| % oppilaista käyttänyt adjektiivia, jota kukaan muu ei käytä | 76 | | 89 | | 97 | |

2.-luokkalaisten aineistossa on yhteensä 464 adjektiivisanetta, joista tyttöjen kirjoitelmissa 64 % ja poikien 36 %. Tyttöjen kirjoitelmissa adjektiivit kattavat vajaat 5 % ja poikien kirjoitelmissa noin 4 % kaikista saneista. Jos suhteutetaan adjektiivien esiintyminen oppilaiden määrään, keskimäärin tyttöjen kirjoitelmissa on 2,6 adjektiivia ja poikien 1,5 adjektiivia. 2.-luokkalaiset käyttävät siis adjektiiveja erittäin niukasti. Viidenes oppilaista käytti kirjoitelmassaan vain yhtä adjektiivia. Toistoa esiintyi hivenen enemmän tyttöjen kirjoitelmissa: yksi adjektiivi esiintyi tyttöjen kirjoitelmissa keskimäärin 3,0 kertaa, poikien kirjoitelmissa 2,6 kertaa.

Kymmenestä eniten adjektiiveja käyttäneestä oppilaasta (4,5 % kaikista 2. luokan oppilaista), seitsemän oli tyttöjä ja kolme poikia. Yhteensä he käyttivät 39 eri adjektiivileksemiä, mikä kattaa 30 % kaikista 2.-luokkalaisten käyttämistä adjektiivileksemeistä. Heidän käyttämänsä adjektiivit kattavat 16 % kaikista 2.-luokkalaisten käyttämistä adjektiivisaneista.

Enimmillään adjektiivisaneita esiintyi 11 kappaletta yhden tytön kirjoitelmassa, joka sisälsi yhteensä 97 sanetta. Hän käytti adjektiiveja *tuhma*, *kiltti* ja *hyvä* yhteensä seitsemän kertaa ja loppuja adjektiiveja kerran. Adjektiivit kattoivat siis hänen kirjoitelmastaan 10 %. Vertailun vuoksi todettakoon, että poikien kirjoitelmien suurin adjektiivimäärä oli 9 kappaletta, joka löytyi kahden pojan aineesta. Toisen pojan aineessa oli 40 sanetta ja toisen 134. 40 sanetta kirjoittanut käytti kirjoitelmassaan kuutta eri

adjektiivilekseemiä, 134 sanetta kirjoittanut ei taas käyttänyt yhtäkään adjektiivia enempää kuin kerran. Tulos on mielenkiintoinen. Yksittäisten kirjoitelmien kohdalla adjektiivisaneiden määrä ei siis välttämättä kerro paljonkaan kirjoitelman pituudesta eikä tasosta, mutta jatkon kannalta olisi mielenkiintoista tutkia, kertooko adjektiivilekseemien määrä niistä jotakin.

2.-luokkalaisten käyttämistä adjektiivilekseemeistä 58 % (75 kpl) esiintyy kirjoitelmissa vain kerran. Tällaisia vain kerran esiintyviä lekseemejä on tyttöjen kirjoitelmissa 66 % ja poikien kirjoitelmissa 34 %. Oppilaista 76 % ainakin kerran sellaista adjektiivilekseemiä, jota kukaan muu ei käytä. Tytöt käyttävät kaikista kirjoitelmissa esiintyneistä adjektiivilekseemeistä 74 %:a ja pojat 49 %:a.

4.-luokkalaisten aineistossa on yhteensä 949 adjektiivisanetta, joista tyttöjen kirjoitelmissa 68 % ja poikien 32 %. Jos suhteutetaan adjektiivien esiintyminen oppilaiden määrään, keskimäärin tyttöjen kirjoitelmissa on 5,5 adjektiivia ja poikien 3,1 adjektiivia. Oppilaista 8 % (17 kpl) käytti kirjoitelmassaan vain yhtä adjektiivia. Näistä oppilaista neljä oli tyttöjä ja loput poikia. Toistoa esiintyi neljännelläkin luokalla enemmän tyttöjen kuin poikien kirjoitelmissa: yhtä adjektiivilekseemiä käytettiin tyttöjen kirjoitelmissa keskimäärin 3,5 kertaa, poikien kirjoitelmissa 2,7 kertaa.

Kymmenestä eniten adjektiiveja käyttäneestä oppilaasta (4,7 % kaikista 4. luokan oppilaista) yhdeksän oli tyttöjä ja vain yksi oli poika. Yhteensä he käyttivät 75 eri adjektiivilekseemiä, mikä kattaa noin 32 % kaikista 4.-luokkalaisten käyttämistä adjektiivilekseemeistä. Heidän käyttämänsä adjektiivit (153 sanetta) kattavat 16 % kaikista 4.-luokkalaisten käyttämistä adjektiivisaneista aivan kuten 2.-luokkalaisillakin.

Enimmillään adjektiiveja esiintyi 18 kappaletta yhden tytön kirjoitelmassa, joka sisälsi 139 sanetta. Adjektiivit kattoivat siis hänen kirjoitelmastaan 13 %. Hän käytti vain kahdeksaa adjektiivilekseemiä, joista adjektiivia *iso* hän käytti jopa kuusi ja adjektiivia *ihana* neljä kertaa. Hänen kirjoitelmansa on arvioitu tasolle skripti, eli adjektiivisaneiden suuresta määrästä huolimatta kirjoitelma ei yllä parhaalle tasolle.

4.-luokkalaisten käyttämistä adjektiivilekseemeistä 57 % (133 kpl) esiintyy kirjoitelmissa vain kerran. Niistä 68 % esiintyy tyttöjen ja 32 % poikien kirjoitelmissa. Oppilaista 89 % (191 kpl) käyttää vähintään kerran jotakin sellaista lekseemiä, jota kukaan muu ei käytä. Tytöt käyttävät kaikista kirjoitelmissa esiintyneistä adjektiivilekseemeistä 80 %:a ja pojat 47 %:a.

6.-luokkalaisten aineistossa on yhteensä 1 022 adjektiivisanetta, joista tyttöjen kirjoitelmissa 64 % ja poikien 36 %. Tyttöjen kirjoitelmissa adjektiivit kattavat noin 6 % ja poikien kirjoitelmissa noin 5 % kaikista saneista. Jos suhteutetaan adjektiivien esiintyminen oppilaiden määrään, keskimäärin tyttöjen kirjoitelmissa on 8,4 adjektiivia ja poikien 4,5. Oppilaista 7 % (11 kpl) käytti kirjoitelmassaan vain yhtä adjektiivia. Näistä oppilaista kolme oli tyttöjä ja loput poikia. Toistoa esiintyi tyttöjen kirjoitelmissa jälleen enemmän kuin poikien kirjoitelmissa: yhtä adjektiivilekseemiä käytettiin tyttöjen kirjoitelmissa keskimäärin 3,4 kertaa, poikien kirjoitelmissa 3,0 kertaa.

Kuudennellakin luokalla kymmenestä eniten adjektiiveja käyttäneestä oppilaasta (6,2 % kaikista 6. luokan oppilaista) yhdeksän oli tyttöjä ja vain yksi oli poika. Yhteensä he käyttivät 81 eri adjektiivilekseemiä, mikä kattaa 33 % kaikista 6.-luokkalaisten käyttämistä adjektiivilekseemeistä. Heidän käyttämänsä adjektiivit (175 sanetta) kattavat 17 % kaikista 6.-luokkalaisten käyttämistä adjektiivisaneista.

Enimmillään adjektiiveja esiintyi 22 kappaletta yhden tytön kirjoitelmassa, joka sisälsi 168 sanetta. Adjektiivit kattoivat siis hänen kirjoitelmastaan 13 %. Hän käytti yhteensä 16 adjektiivilekseemiä, joista useammin kuin kerran adjektiiveja *kiva*, *kaunis*, *jännittävä*, *suuri* ja *uskomaton*. Hänen kirjoitelmansa on arvioitu tasolle ei skeemaa.

6.-luokkalaisten käyttämistä adjektiivilekseemeistä 57 % (142 kpl) esiintyy kirjoitelmissa vain kerran. Niistä 66 % esiintyy tyttöjen ja 34 % poikien kirjoitelmissa. Oppilaista 97 % käyttää vähintään kerran jotakin sellaista adjektiivilekseemiä, jota kukaan muu ei käytä. Tytöt käyttävät kaikista kirjoitelmissa esiintyneistä adjektiivilekseemeistä 79 %:a ja pojat 49 %:a.

Tulosten perusteella voi todeta, että adjektiivien käyttö monipuolistuu huomattavasti ylemmillä luokilla. Tyttöjen kirjoitelmissa adjektiivisaneiden määrä myös selvästi lisääntyy. Poikienkin kirjoitelmien keskimääräinen adjektiivisaneiden määrä kasvaa, mutta ei lainkaan yhtä paljon kuin tyttöillä. Sen sijaan adjektiivisaneiden osuus kaikista saneista kasvaa sekä tyttöjen että poikien kirjoitelmissa selvästi.

Adjektiivien kumuloituvaa frekvenssiä kuvastaa hyvin se, että joka luokalla jo kymmenen eniten adjektiiveja käyttänyttä oppilasta on käyttänyt lähes kolmasosaa kaikista luokan kirjoitelmissa esiintyneistä adjektiivilekseemeistä. Näiden oppilaiden käyttämien adjektiivilekseemien osuus kaikista luokan adjektiivilekseemeistä kasvaa hivenen ylemmillä luokilla, mikä saattaa kertoa siitä, että erot adjektiivien käytössä kasvavat siirryttäessä ylemmille vuosiluokille.

2., 4. ja 6. luokkien kymmenen frekventeimmän adjektiivin lista (taulukko 12) näyttää sisältävän lähes samat adjektiivit, mutta kaikille yhteisiä adjektiiveja ovatkin vain *oma*, *kiva*, *iso*, *hyvä* ja *paras*. Nämä viisi adjektiivia ovat myös koko aineiston viisi frekventeintä adjektiivilekseemiä (kuvio 14). Vaikka kaikille yhteisiä adjektiiveja löytyy listoilta vain viisi, ainoastaan kaksi adjektiivia, *viimeinen* ja *lämmin*, esiintyy pelkästään yhdellä listalla.

Taulukko 12. 2.-, 4.- ja 6. -luokkalaisten tyttöjen ja poikien 10 eniten käyttämää adjektiivia.

| 2. luokka | | 4. luokka | | 6. luokka | |
|-----------|------------------|-----------|-----------|---------------|-----------|
| tytöt | pojat | tytöt | pojat | tytöt | pojat |
| oma | kiva | oma | iso | hyvä | hieno |
| kiva | uusi | ihana | hieno | paras | kiva |
| pieni | hauska | kiva | hyvä | iso | hyvä |
| iso | paras | iso | hauska | pitkä | iso |
| hauska | hyvä | hyvä | uusi | tosi | paras |
| hieno | iso | pieni | kiva | ihana | oma |
| hyvä | oma | uusi | seuraava | kiva | pitkä |
| uusi | pieni | hauska | ihana | oma | sellainen |
| pitkä | pitkä | paras | sellainen | lämmin | seuraava |
| paras | viimeinen | seuraava | oma | seuraava | tosi |
| | | suuri | paras | | uusi |
| | | | pieni | | |

4.- ja 6.-luokkalaisten listoilla on *seuraava*, jota ei pienempien oppilaiden listalta löydy. Se on hieman korkeammalla sijalla poikien kuin tyttöjen listoilla. *hauska* on erittäin käytetty adjektiivi 2. ja 4. luokalla, mutta 6. luokan listoista sitä ei löydy. *sellainen* esiintyy vain 4. ja 6. luokan poikien listoilla. *uusi* on kaikkien muiden paitsi 6.-luokkalaisten tyttöjen listalla. Pienistä eroista huolimatta yleisimpien adjektiivien vertailu osoittaa, että loppujen lopuksi alakouluikäiset, niin tytöt kuin pojatkin, käyttävät enimmäkseen samoja perusadjektiiveja.

Lopuksi luodaan silmäys siihen, minkälaisia adjektiivilekseemejä tytöt käyttävät, mutta pojat eivät ja päinvastoin. 2.-luokkalaisten aineistossa vähintään kaksi kertaa esiintyvät lekseemit (55 lekseemiä, 2- sanetta, yhteensä 389 sanetta) kattavat noin 84 % koko luokan aineiston lekseemeistä. Vähintään kaksi kertaa esiintyneiden adjektiivilekseemien joukossa on 15 sellaista lekseemiä, jota ovat käyttäneet vain tytöt. Nämä lekseemit ovat *harmaa*, *huono*, *ikinen* (kollokaatista *joka ikinen*), *ikioma*, *kaikenlainen*, *kaunis*, *kiltti*, *lämmin*, *musta*, *onnellinen*, *outo*, *tuhma*, *tuttu*, *valkoinen* ja *värinen*. Vain poikien kirjoitelmissa esiintyneitä lekseemejä oli kuusi: *allerginen*, *dorka*, *ilmainen*, *kullattu*, *kunnon* ja *viimeinen*. Vain kerran käytettyjä adjektiiveja on 2.-luokkalaisten aineistossa yhteensä 75, joista 50 löytyy tyttöjen ja

25 poikien aineistosta. Vain kerran käytetyt adjektiivit kattavat tyttöjen käyttämistä adjektiiveista 17 % ja poikien käyttämistä adjektiiveista 15 %.

6.-luokkalaisten aineistossa vähintään kaksi kertaa esiintyvät lekseemit (104 lekseemiä, 2-52 sanetta, yhteensä 877 sanetta) kattavat noin 86 % kaikista 6.-luokkalaisten kirjoitelmien adjektiivisaneista. Näiden 104 frekventeimmän lekseemin joukossa on 31 sellaista lekseemiä, jota ovat käyttäneet vain tytöt. Nämä lekseemit ovat *antiikinaikainen, hirveä, ihastuttava, ikuinen, ikimuistoinen, innokas, kesäinen, kiltti, kirkas, komea, korkea, kylmä, köyhä, lätsäpäinen, mitätön, musta, nuori, pilvetön, puhdas, rauhallinen, ruskea, samanlainen, sopiva, stylekäs, suloinen, todellinen, uskomaton, valkoinen, vanha, viereinen ja viileä*. Vain poikien käyttämiä adjektiiveja oli sadan yleisimmän adjektiivilekseemin joukossa neljä: *huipputehokas, maaliton, selvä ja tyhjä*. Vain kerran käytettyjä adjektiiveja on 6.-luokkalaisten aineistossa yhteensä 142, joista 94 löytyy tyttöjen ja 48 poikien kirjoitelmista. Kerran käytetyt adjektiivit kattavat tyttöjen käyttämistä adjektiiveista 14 % ja poikien käyttämistä adjektiiveista 13 %.

Koko aineistossa vähintään viisi kertaa esiintyvät lekseemit (116 lekseemiä, 5-207 sanetta, yhteensä 3 530 sanetta) kattavat noin 82 % kaikista adjektiivisaneista. Niiden joukossa on yhdeksän vain tyttöjen käyttämää lekseemiä ja yksi vain poikien käyttämä lekseemi. Vain tyttöjen käyttämät lekseemit olivat *hirveä, innokas, minkälainen, puhdas, ruskea, sopiva, syvä, turvallinen ja viileä*, vain poikien käyttämä adjektiivi taas oli *helppo*. Vähintään kaksi kertaa käytettyjen lekseemien joukossa (yhteensä 278 kappaletta) on 76 sellaista lekseemiä, joita ovat käyttäneet vain tytöt ja 18 sellaista lekseemiä, jota ovat käyttäneet vain pojat. Kerran käytettyjä lekseemejä on koko aineistossa yhteensä 336 kappaletta, joista tyttöjen käyttämiä on 203 ja poikien käyttämiä 133 kappaletta. Koko aineistossa on siis yhteensä 279 sellaista adjektiivilekseemiä, jota pojat eivät käytä ja 151 sellaista adjektiivilekseemiä, jota tytöt eivät käytä. Kirjoittajamääriin suhteutettuna ”tyttömäisiä” adjektiiveja käyttää noin joka toinen tyttö ja ”poikamaisia” adjektiiveja noin joka neljäs poika.

Edellä olevien tietojen valossa tyttöjen adjektiivien käyttö vaikuttaisi huomattavasti runsaammalta ja monipuolisemmalta kuin poikien. Myös värit näyttävät kuuluvan enemmän tyttöjen kuin poikien sanastoon. Kuisma (2011) on tutkinut Koulukorpus-aineistosta ominaisuudennimien eli käytännössä adjektiivien käytön kehittymistä muun muassa ilmausten määrän, sanapituuden, sanaston diversiteetin ja semanttisen laadun suhteen. Hänenkin

tulostensa perusteella tytöt käyttävät keskimäärin enemmän adjektiiveja kirjoittajaa kohden kuin pojat myös sanemääriin suhteutettuna.

Opettajan sanastokirjan (1970: 43) mukaan kielitaidon kehittyessä adjektiivien käyttö lisääntyy. Voidaan siis todeta, että tytöt ovat adjektiivien käytön perusteella kehittyneempiä kielenkäyttäjiä kuin pojat. Aineen tason kanssa adjektiivien määrä ei kuitenkaan näytä korreloivan (ks. kuviot 9 ja 10 s. 25). Tätä selittää se, että aineiden tasoluokituksissa on otettu aina huomioon oppilaiden ikä. Ylempien luokkien kertomukset ovat siis hyvin erilaisia kuin alemmien luokkien kertomukset ja sama pätee muunkin tasoihin aineisiin. Tästä syystä eritasoisia aineita pitäisikin loppujen lopuksi verrata vain vuosiluokan sisällä, sillä kaikkien vuosiluokkien niputtaminen yhteen aiheuttaa edellä mainitun ristiriidan.

3.2.4 Adverbit

Adverbit ovat taipumattomia tai vajaaparadigmaisia, esimerkiksi aikaa, paikkaa, tapaa, ja määrää ilmaisevia sanoja. Koulukorpus-aineiston adverbien peruskoodauksessa on joitakin epäyhtenäisyyksiä siinä, onko jokin adverbialli palautettu lekseeminä sanajuureen vai luokiteltu kiteytyneeksi adverbiksi, mikä saattaa jossakin kohdin vaikuttaa tuloksiin.

Koivunen (2012) on pro gradu -työssään tutkinut alakouluikäisten adverbien käyttöä yksityiskohtaisesti. Hänen tulostensa mukaan tytöt käyttävät alemmilla luokilla enemmän adverbejä kuin pojat, mutta erot tasoittuvat kouluvuosien myötä. Ajan, paikan ja tavan adverbien käyttö lisääntyy ja määrän ja proadverbien käyttö vähenee. Tässä tutkimuksessa adverbien käyttöön luodaan vain pikainen silmäys.

Koko aineiston kymmenen eniten käytettyä adverbiä kattavat lähes 40 % adverbeistä ja 30 yleisimmän adverbien osuus on lähes 60 %. 30 yleisimmän adverbien suhteelliset osuudet on esitetty kuviossa 15. Koko aineiston yleisimmät adverbit esitetään liitteessä 6.

Adverbien käyttö näyttää olevan alemmilla vuosiluokilla jonkin verran ylempiä vuosiluokkia vähäisempää (ks. kuvio 7 s. 22). Siinä on nähtävissä myös jonkinlaista kehitystä puhekielisyydestä yleiskielisyyden suuntaan, kun sitä tarkastellaan luokkakohtaisesti. Alemmilla luokilla käytetään suhteellisen paljon adverbiä *tosi* erityisesti vahvistussanana (esimerkit 8 ja 9), kun taas ylemmillä luokilla sitä käytetään suhteellisesti vähemmän. Samalla adverbien *todella*, *hyvin* ja *erittäin* käyttö lisääntyy.

8) Se on nykyään tosisuosittu lomakohde. (taajama 2P128)

9) Se olisi tosi tosi kiva päivä. (kaupunki 2T46)

2.-luokkalaiset käyttävät *tosi*-adverbiä vahvistussanana 75 %:ssa tapauksista, kun taas 6.-luokkalaisilla *tosi* kattaa alle puolet vahvistussanoista. Ylemmillä luokilla *tosi* korvataan adverbeilla *hyvin*, *todella* ja *erittäin*. 2.-luokkalaiset eivät ole käyttäneet *erittäin*-adverbiä vahvistussanana kertaakaan.

Nuoremmat kirjoittajat käyttävät myös *kerran*-adverbiä huomattavan paljon vanhempiin oppilaisiin verrattuna (kuvio 15). *kerran*-sanaa on käytetty 2.-luokkalaisten aineistossa yhteensä 35 kertaa, joista 25 kertaa sitä on käytetty aloittamalla kirjoitelma saduista tutulla *olipa kerran* -fraasilla (esimerkki 10). Muutenkin *kerran*-sanon käyttö näyttää yleensä liittyvän tarinamaiseen kerrontaan (esimerkki 11). 6.-luokkalaisista vain viisi aloittaa kirjoitelmansa *olipa kerran* -fraasilla.

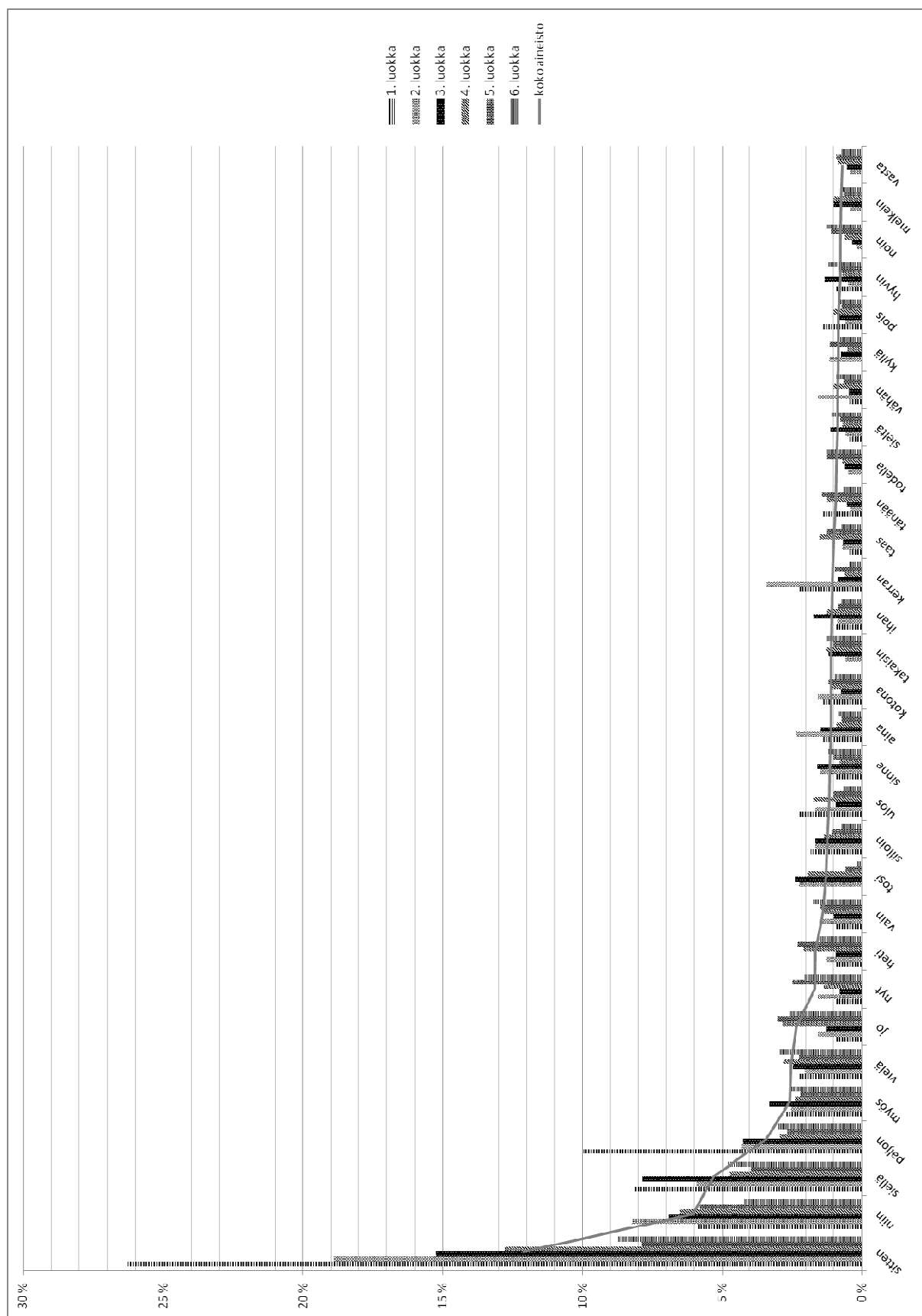
10) Olipa kerran tyttö joka piti hevosista paljon. (kaupunki 2T35)

11) Kerran minä ja minun kaverini oltiin metsässä sienessä. (kaupunki 2T32)

sitten-adverbin käyttö on selvästi runsaampaa alemmilla kuin ylemmillä vuosiluokilla. Tähän on syynä todennäköisesti nuorempien kirjoittajien luettelomainen tyyli (*sitten*/*ja sitten*: esimerkki 12).

12) Jasitten se oli kotona joka päivä. Ja sitten mulla olisi hauska joka päivä. (kaupunki 2T2)

Vanhemmat ja taitavammat oppilaat hallitsevat muunlaisia, sujuvampia keinoja ilmaista siirtymiä kirjoitelmissaan. Myös Opettajan sanastokirjassa (1970: 42) esitetään, että *sitten*-adverbiä käytetään kuvaamaan yksinkertaista ajallista järjestystä ja että sen käyttö kuvastaa kielen kehittymättömyyttä. Väite käy hyvin yksin tämän tutkimuksen tulosten kanssa. Opettajan sanastokirjan mukaan partikkelien (sisältää adverbit ja muut taipumattomat sanat) runsaus tekstissä on myös yhteydessä kielen kehittyneisyyteen (mts.). Ainakin adverbien osuus saneista näyttää tämän aineiston valossa kasvavan ylemmillä vuosiluokilla.



Kuvio 15. Koko aineiston 30 yleisimmän adverbin suhteelliset osuudet.

3.2.5 Pronominit

Pronominilekseemin määrittely ei ole aivan yksioikoista. Koulukorpuksen peruskoodauksessa on määritelty omaksi lekseemikseen sekä pronominin epäsäännölliset taivutusmuodot että vain nominatiivimuotoinen pronomini. Vaihtelua on näkynyt myös siinä, kuinka käsitellään muotoja, jotka voisi määritellä kiteytyneiksi. Tässä tutkimuksessa saneita käsitellään perinteisen määritelmän mukaan siten, että kaikki yhden sanan taivutusmuodot ovat saman lekseemin konkretisoitumia joitakin leksikaalistuneita poikkeuksia lukuun ottamatta. Siksi on loogista, että esimerkiksi saneita *siitä* ja *sen* edustaa lekseemi *se* sanavartalon epäsäännöllisyydestä huolimatta. Toisaalta on otettava huomioon myös sanan merkitys lauseessa. Esimerkiksi *tuolla* voi käyttöyhteydestään riippuen olla joko paikkaa ilmaiseva adverbi *tuolla* tai demonstratiivipronominin *tuo* adessiivimuoto.

Alakouluikäisten pronominiinien käyttöä on pro gradu -työssään semanttisesti tutkinut myös Palonen (tekeillä), joten tässä työssä pronominiinien tarkastelu painottuu yleisesti lekseemitasolle. Koko aineisto sisältää vajaat 8500 pronominia. Pronominilekseemien määrä on 1.-luokkalaisilla 18. Jo 2.-luokkalaisilla pronominilekseemien määrä nousee yli 30:n ja 6.-luokkalaisilla se on 36. Kuten muuhunkin sanastoon, myös pronominiinien käyttöön vaikuttaa aineen aihe. Koska aihe on hyvin omakohtainen, unelmieni päivä, aine on kirjoitettu yleensä minä-muodossa, mikä lisää *minä*-pronominin käyttöä. Toinen syy runsaaseen *minä*-pronominin käyttöön erityisesti alemmilla luokilla on pronominiinien ”ylimääräinen” käyttö (esimerkki 13).

13) Minä saan oman lemmikin. minä olen aina unelmoinut lemmikistä. Minä saan hamsterin. (kaupunki 2T1)

Vanhemmat oppilaat ovat jo oppineet, että persoonapronomineja voi jättää pois, sillä persoonamuoto näkyy verbin päätteessä.

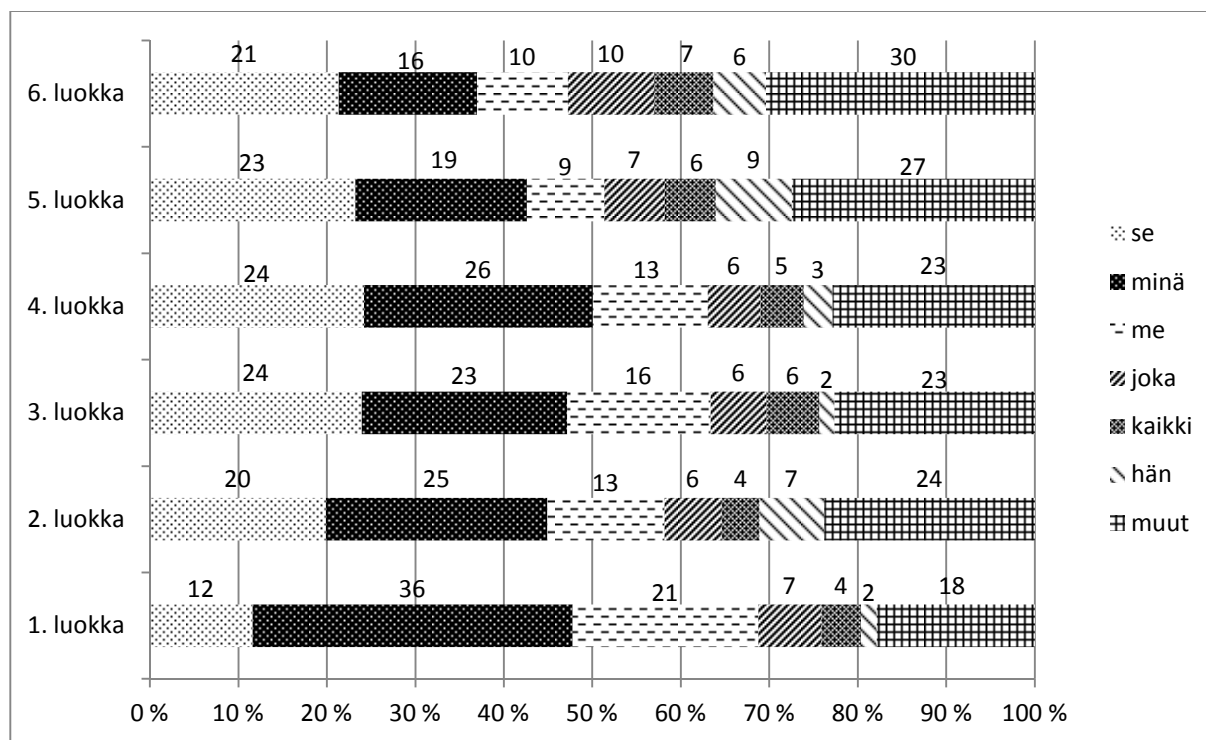
Myös *me*-pronominia ”ylimääräistä” käyttöä esiintyy samalla tavalla kuin *minä*-pronomin, minkä lisäksi monet nuoret oppilaat tarvitsevat sitä runsaasti siksi, että he käyttävät puhekielistä passiivimuotoa kirjoittaessaan me-muodossa (esimerkit 14, 15 ja 16).

14) Uutenavuotena me mentiin sisälle ja me alettiin pelata ja sitten me mentiin ulos ja -- (maalais 1T47)

15) Kun me oltiin Linanmäellä Me saatiin paljon herkkua siellä. (kaupunki 3P131)

16) Ja me hypittiin trampoliinilla. (taajama 3T2)

Seuraavaksi tarkastellaan pronominiin suhteellisia osuuksia eri vuosiluokkien kirjoitelmissa (kuvio 16). Kuuden frekventimmän pronominin osuudet näytetään erikseen ja loput on niputettu yhteen.



Kuvio 16. Pronominilekseemien suhteelliset osuudet luokittain.

Aineiston kuusi yleisintä pronominia ovat *se*, *minä*, *me*, *joka*, *kaikki* ja *hän*. *minä*-pronominin käytön väheneminen 1. luokan jälkeen on yksi näkyvimpiä muutoksia pronominiin käytössä. Myös *me*-pronominin käyttö vähenee. Sen sijaan *se*-pronominin käyttö lisääntyy ja vakiintuu heti 2.-luokalta lähtien. Tämä voi olla seurausta substantiivien toiston vähentämisestä pronominiin käytön avulla. *hän*-pronomini on ainoa, jonka suhteellisen osuuden muutoksissa ei näy selkeää trendiä. Kuusi yleisintä pronominia kattavat 1.-luokkalaisten käyttämistä pronomineista yli 80 % ja 2.-luokkalaisten pronomineista lähes 76 %. 6.-luokkalaisten käyttämistä pronomineista ne kattavat enää 70 %.

Opettajan sanastokirjassa (1970: 43) esitetään, että kehittyneemmät kielenkäyttäjät käyttävät enemmän pronomineja *hän*, *se*, *te*, *ne* ja *he*, kun taas kehittymättömämmät kielenkäyttäjät käyttävät enemmän pronomineja *minä* ja *me*. Relatiivipronominiin käyttö kertoo myös kehittyneemmistä kirjoitustaidoista (Matilainen 1989: 117). Tämän tutkimuksen aineistossa 2.-luokkalailla *minä* ja *me* kattaa 38 % pronomineista, 4.-luokkalailla 39 % ja 6.-

luokkalaisilla 26 %. Epäilemättä minäkeskeisellä aiheella on ollut osuutta runsaaseen omakohtaisten pronominin käyttöön vielä 4. luokallakin, sillä Opettajan sanastokirjan esittämien periaatteiden valossa 4.-luokkalaisten pronominin käyttö näyttää *se*-pronominia lukuun ottamatta 2.-luokkalaisten pronominin käyttöä kehittymättömämmältä (taulukko 13).

Taulukko 13. Eräiden pronominin sanemäärät per tuhat pronominisanetta.

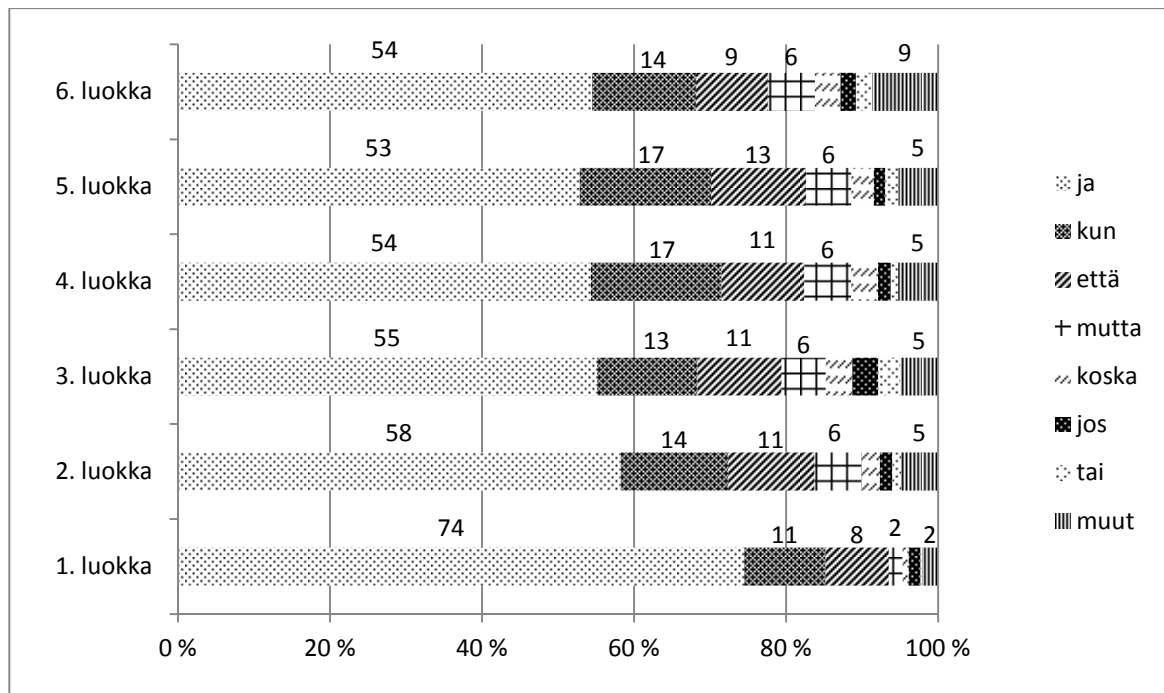
| | 2. luokka | 4. luokka | 6. luokka |
|-------------|-----------|-----------|-----------|
| <i>minä</i> | 250 | 258 | 156 |
| <i>me</i> | 132 | 130 | 103 |
| <i>hän</i> | 73 | 33 | 59 |
| <i>se</i> | 199 | 242 | 214 |
| <i>joka</i> | 65 | 59 | 97 |

Pronominien runsautta pidetään kehittyneen kielen merkinä, mutta Koulukorpus-aineisto ei siinäkään suhteessa noudattele Opettajan sanastokirjassa esitettyjä tuloksia, sillä ylemmien vuosiluokkien kirjoitelmissa ei ole käytetty enempää erilaisia pronominileksemejä kuin alemmien vuosiluokkien kirjoitelmissa ja pronominisaneiden suhteellinen osuus on jopa pienempi ylemmien kuin alemmien vuosiluokkien kirjoitelmissa. Laadullinen muutos pronominilekseemien käytössä kuitenkin näkyy siinä, että ylemmillä luokilla käytetään pronomineja monipuolisemmin kuin alemmilla luokilla.

3.2.6 Konjunktiot

Konjunktioiden tehtävänä on sitoa saman virkkeen lauseita tai lauseen osia toisiinsa. Eri kielioppien määritelmät siitä, mitkä sanat ovat konjunktioita ja mitkä esimerkiksi adverbejä saattavat hieman erota toisistaan. Esimerkiksi proadverbi *joten* voitaisiin määritellä adverbiksi, mutta Koulukorpuksen alkuperäisessä koodauksessa se on määritetty konjunktioksi. Tässä tutkimuksessa noudatetaan alkuperäisen koodauksen linjaa. Konjunktiot jaetaan perinteisesti rinnastuskonjunktioihin, joita ovat esimerkiksi *ja*, *tai*, *mutta*, *sekä* ja *eli*, sekä alistuskonjunktioihin, joita ovat esimerkiksi *kun*, *kuin*, *jos*, *koska* ja *vaikka* (ks. esim. Leino 1987).

Konjunktioiden käyttö kertoo muun muassa siitä, kuinka lauseita sidostetaan toisiinsa. Koska on jo todettu (ks. Pajunen 2012), että nuorille kirjoittajille ominaista on esimerkiksi lauseiden ketjuttaminen tavalla, jota kehittyneet kirjoittajat eivät tee, alakoulun oppilaiden konjunktioiden käytön tarkastelu ja vertailu muuhun aineistoon on relevanttia.



Kuvio 17. Konjunktioiden käyttö luokittain.

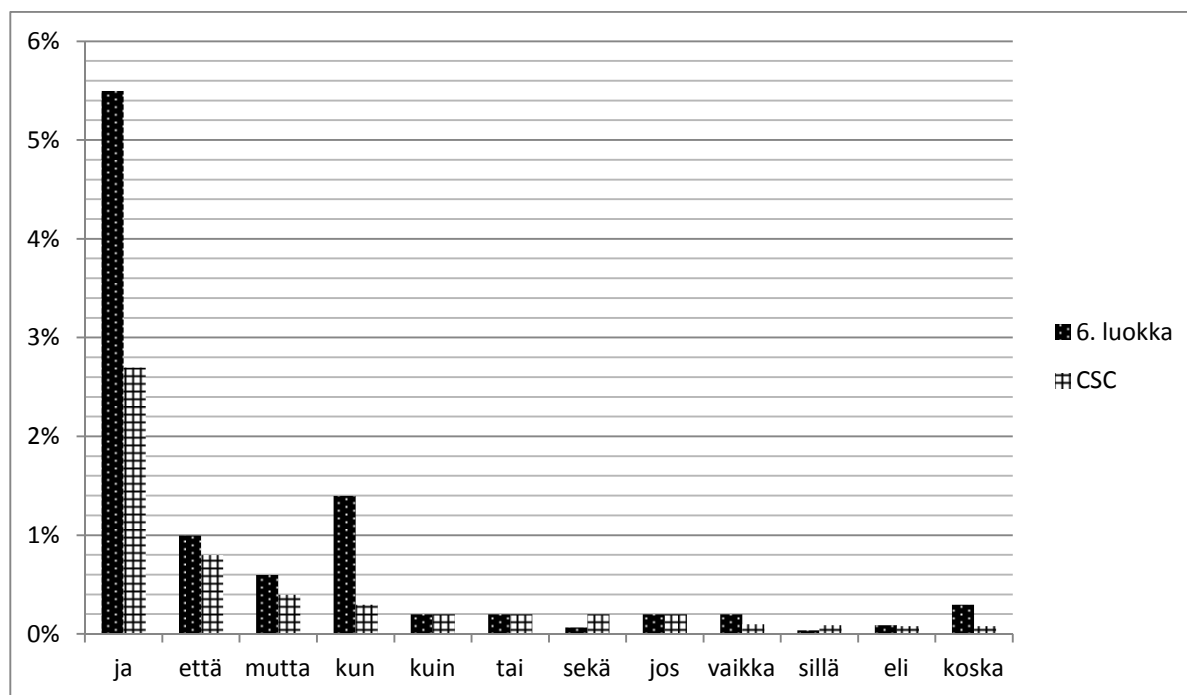
Yleisimmin käytetty konjunktio on *ja*, jonka osuus on kaikilla luokilla yli puolet kaikista käytetyistä konjunktioista (kuvio 17). Seuraavaksi eniten käytetään *kun* ja *että* -konjunktioita, ja niiden käyttö on huipussaan 5. luokalla. 1.-luokkalaiset käyttävät *ja*-konjunktia huomattavan paljon, lähes kolme neljästä konjunktiosta on *ja*-konjunktio. 2.-luokkalaisilla *ja*-konjunktion käyttö jää alle 60 %:n, ja vanhemmat oppilaat käyttävät sitä vielä hieman vähemmän.

ja-konjunktia käytetään alemmilla luokilla ennemminkin konnektiivina, ja sen runsaus johtuu nuorten kirjoittajien tavasta ketjuttaa lauseita sen avulla (esimerkki 17).

17) sitten isommat pojat tulivat ja kiusasiv pienempiä ja sitten minä menin kertomaan äitille ja äiti meni hakemaan muut pojat ja äiti tarjosi meille kaikille jäätelöä ja äiti antoi minulle 15 euroa. (kaupunki 2P155)

1. -luokan jälkeen konjunktioiden käyttö ei näyttäisi muuttuvan merkittävästi. Suurin vaihtelu tapahtuu *kun* ja *että* -konjunktioiden osalta. Merkittävintä niiden käytön lisäksi on se, että 6.-luokkalaisten konjunktioiden käyttö on selvästi monipuolistumassa: muiden kuin aineiston seitsemän yleisimmän konjunktion osuus on 9 %, kun se muilla luokilla on parhaimmillaankin vain 5 %.

Konjunktoiden osuutta Koulukorpus-aineistossa on mahdollista verrata muuhunkin aineistoon (kuvio 18), sillä kirjoitelmien aihe ei välttämättä vaikuta konjunktoiden käyttöön yhtä ratkaisevasti kuin esimerkiksi substantiiveihin ja adjektiiveihin. CSC:n sanomalehtikielen taajuussanastossa yleisimmät konjunktiot ovat *ja*, *että*, *mutta*, *kun*, *kuin*, *tai*, *sekä*, *jos* ja *vaikka*. Nämä konjunktiot kuuluvat taajuussanaston sadan frekventeimmän lekseemin joukkoon. *koska* on sijalla 106. *koska* ja *sillä* ovat yleensä vaihtoehtoiset, ja *sillä* saatetaan valita usein siksi, että se kuulostaa pehmeämmältä. Sanomalehtikieleen verrattuna 6.-luokkalaiset käyttävät edelleen hyvin paljon *ja*-konjunktia: Heillä sen osuus saneista on noin 5,5 %, kun sanomalehtikielessä *ja*-konjunktion – sanomalehtikielessäkin toiseksi yleisimmän lekseemin – osuus on vain noin 2,7 %. Muiden yleisimpien konjunktoiden osalta erot eivät prosenttiyksiköissä ole aivan yhtä suuret: *kun*-konjunktio kattaa 6.-luokkalaisilla 1,4 % saneista ja sanomalehtikielessä 0,3 %, *mutta*, *että* ja *kuin* -konjunktoiden osuus saneista on molemmissa aineistoissa lähes yhtä suuri. *sekä*-konjunktia käytetään 6.-luokkalaisten kirjoitelmissa huomattavasti vähemmän kuin sanomalehtikielessä. On mahdollista, että 6.-luokkalaiset eivät vielä osaa käyttää *sekä*-konjunktia sujuvasti *ja*-konjunktion sijasta.



Kuvio 18. CSC:n Suomen sanomalehtikielen taajuussanaston (2004) ja 6.-luokkalaisten käyttämien konjunktoiden osuudet saneista.

Verrattuna sanomalehtikieleen *kun*-konjunktia käytetään runsaasti koululaisten kirjoitelmissa. *kun*-konjunktion suurta frekvenssiä selittää osin aineen aihe, sillä tarinat

unelmien päivästä rakentuvat usein kronologisesti; tapahtumasta tai vuorokaudenajasta toiseen siirrytään usein virkkeellä, jonka aloittaa *kun*-konjunktioilla alkava sivulause (esimerkit 18 ja 19).

18) Kun olisin syönyt lähtisin hotellille. (kaupunki 6T10)

19) Kun kello on yhdeksän illalla alan katsoa jotain elokuvaa. (kaupunki 6P83)

Lauseenvastikkeita kuudesluokkalaisetkin käyttävät vielä varsin vähän, kun otetaan huomioon temporaalisten sivulauseiden suuri määrä. Kaikkien lauseenvastikkeiden osuus on noin 2 % ja temporaalisten lauseenvastikkeiden noin 0,6 % kaikista verbisaneista. Tämäkin saattaa olla yksi syy *kun*-konjunktion runsaaseen käyttöön. *kun*-konjunktion käyttöön liittyy myös puhekielelle tuttu ilmiö: *kun*-konjunktioita käytetään tilanteissa, joissa oikeakielisyysääntöjen mukaan *joka*-relatiivipronomini olisi korrektimpi vaihtoehto (esimerkki 20).

20) Unelmieni päivä alkaisi kesäisenä aamuna kun linnut laulaisivat ja saisin nukua pitkään (kaupunki 6T8)

Puhekielen vaikutus kirjoitelmien lauserakenteisiin on nähtävissä monin tavoin, korrektin kielenkäytön näkökulmasta myös niin sanottuna ylimääräisenä *kun*-konjunktion käyttöä (esimerkit 21–23).

21) Kun oli lauantai niin menime Tampereel (taajama 6P118)

22) Unelmien päivä oli kun saimme uuden koiran. (kaupunki 6T21)

23) Minun unelma päiväni oli kun sain mopon (maalais 6P148)

Esimerkkivirkkeitä voisi hyvin tiivistää siten, että niissä olisi vain yksi lause ilman, että merkitys muuttuisi. Sanomalehtitekstissä pyritään epäilemättä mahdollisimman tiiviiseen ilmaisuun, mikä tarkoittaa esimerkiksi turhien sivulauseiden pois jättämistä.

Temporaalista *kun*-alkuista sivulauseetta saatetaan käyttää silloinkin, kun referatiivinen *että*-sivulause voisi olla sopivampi lauseen merkityksen kannalta (esimerkit 24 ja 25). Myös syytä ilmaisevissa sivulauseissa käytetään *kun*-konjunktioita (esimerkki 26).

24) En katsonut peliä mutta kuulin kun joku teki maalin. (kaupunki 6P88)

25) Voi harmi kun he eivät olleet huomanneet autoa. (kaupunki 6T22)

26) Avaruusaluksessa olisi kiva olla kun siellä ei ole painovoimaa ollenkaan jos se on avaruudessa eikä vaikka kuussa. (kaupunki 6P92)

kun-konjunktia käytetään myös *kuin*-konjunktion asemesta vertailua tehtäessä, mikä on varsin tavallista puhekielessäkin (esimerkki 27).

27) Haluaisin kävellä ja hyppiä kuussa koska siellä painovoimaa ei ole niin paljon kun maassa. (kaupunki 6P92)

Esimerkkien ja *kun*-konjunktion frekvenssin perusteella koululaiset käyttävät *kun*-konjunktia myös kielenhuollon näkökulmasta epäkorrekteilla tavoilla. Puhekielen vaikutus on nähtävissä monimutkaisissa lauserakenteissa, joita ei esimerkiksi sanomalehtikielessä pitäisi esiintyä, mikäli kaikki lehteen menevä, myös yleisönosaston kirjoitukset, käyvät läpi kielenhuollon ennen julkaisua.

3.2.7 Numeraalit

Numeraalien määritelmä on hyvä ottaa esille. Perinteisen määritelmän mukaan numeraaleja ovat lukusanat, jotka jaetaan tarkemmin peruslukuihin eli kardinaalilukuihin ja järjestys- eli ordinaalilukuihin, joita käytetään lauseessa sekä substantiivin että adjektiivin tavoin. Iso suomen kielioppi (2004) sen sijaan esittää, että vain selkeästi lukumäärää ilmaisevat lukusanat olisivat numeraaleja. Sen sijaan esimerkiksi kellonajat olisivat substantiiveja ja järjestysluvut adjektiiveja. Ison suomen kieliopin kannan seuraaminen merkitsisi sitä, että adjektiivien ja substantiivien osuus tuloksissa kasvaisi ja numeraalien osuus laskisi jonkin verran. Koulukorpuksen peruskoodaus kuitenkin noudattaa perinteistä linjaa, jota noudatetaan edelleen kouluopetuksessakin, eikä tässä tutkimuksessa siis jaotella lukusanoja Ison suomen kieliopin linjan mukaan.

Kielenhuollon näkökulmasta kymmenen ja sitä pienemmät kokonaisluvut olisi suositeltavaa kirjoittaa kirjaimin. Mikäli aineissa esiintyy runsaasti numeroita, voi se kertoa muun muassa siitä, että tätä ohjetta ei vielä ole sisäistetty. Kertooko suhteellisen runsas numeroiden käyttö yleensäkin jotakin kirjoitustaidosta? Numeraalien osuus saneista kasvaa selvästi ylemmillä vuosiluokilla, joskin 4.-luokkalaisten poikien kirjoitelmissa numeraalien

osuus on huomattavan korkea kaikkiin muihin ryhmiin verrattuna (taulukko 14). 1. vuosiluokan jälkeen tytöt käyttävät numeraaleja saneisiin suhteutettuna aina vähemmän kuin pojat. 4. luokalla ero on huomattavan suuri, sillä tyttöjen käyttämien numeraalien osuus noudattaa melko tasaista linjaa, mutta poikien kirjoitelmissa numeroiden osuudessa tapahtuu huomattava hypähdys.

Taulukko 14. Numeraalien osuus saneista tyttöjen, poikien ja koko vuosiluokan kirjoitelmissa.

| | 1. luokka | 2. luokka | 3. luokka | 4. luokka | 5. luokka | 6. luokka |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Tytöt (%) | 0,9 | 0,8 | 1,6 | 1,6 | 1,6 | 1,9 |
| Pojat (%) | 0,9 | 2,0 | 2,1 | 3,7 | 2,5 | 2,4 |
| Kaikki | 0,9 | 1,2 | 2,0 | 2,4 | 2,0 | 2,1 |

Numeroiden käsittelyn aineistossa voi tehdä eri tavoin. Numerot voidaan jättää sivuun tarkastelusta (ks. esim. Jaakola 2004, Niemikorpi 1991), mutta numeroiden tarkastelu Koulukorpuksen kaltaisessa aineistossa saattaa olla aiheellista. Tämä tutkimuksen kohdalla onkin päädytty jälkimmäiseen vaihtoehtoon, sillä runsas numeroiden käyttö saattaa kertoa jotakin lapsen kirjoitustaidosta. Toisin kuin esimerkiksi oppikirjat, kirjoitelmat eivät juurikaan sisällä numeroluetteloja, vaan numerot ovat aina oleellinen osa tekstiä ja enimmäkseen niitä on luokiteltiin melko vähän.

Numeraalit, numerot mukaan lukien, kattavat koko aineiston saneista 2,0 %. Pelkkien numeroiden osuus kaikista saneista on 0,6 %. Eniten numeroita käyttävät 4.-luokkalaiset, joiden kirjoitelmissa niiden osuus on noin 1 % kaikista saneista. 6.-luokkalaisilla numeroiden osuus saneista on 0,9 %, 5.-luokkalaisilla 0,8 %, 3.-luokkalaisilla 0,7 % ja 1. ja 2.-luokkalaisilla 0,4 %. Numeroiden joukkoon ei ole otettu mukaan yhdyssanan määriteosia.

Numeroilla ilmaistaan esimerkiksi määrää, kellonaikoja, rahasummaa, pelituloksia, pelaajien numeroita, järjestystä tai ikää (esimerkit 28–34). Numeroita käytetään välillä silloinkin, kun kielenhuoltosäännösten mukaan lukusana olisi korrektimpaa kirjoittaa kirjaimin.

28) Sitten ostan 200 lentokonetta, 600 autoa, 100 laivaa ja kaikki maailman karkit, purkat, pullat ja jätskit.

(taajama 3T13)

29) Lähden kello 13.09 ja palaan 17.11. (kaupunki 4P116)

30) Saisin syntymäpäivälahjaksi 200 euroa. (taajama 2P196)

31) Tappara voitti sen pelin 10-0. (kaupunki 4P121)

32) Manun kokoonpano olisi seuraava VanderSar #19, Silvestre #21, Feridinand #5 --. (kaupunki 6T16)

33) 1. erässä ei tule maaleja mutta 2. erässä teemme maalin bostonin verkkoon. (kaupunki 6P98)

34) Juuso on 10vuotta vanha. (taajama 3T15)

Numeroiden käytöllä saattaa olla yhteys kirjoitustaitoon. Erityisen paljon numeroita käytetään 3.-luokkalaisten skripteissä, 4.-luokkalaisten listoissa ja 5.-luokkalaisten ei skeemaa -tyyppisissä kirjoitelmissa (taulukko 15).

Taulukko 15. Numeroiden osuus (%) saneista 3.–6. luokan kirjoitelmissa.

| | kertomus | skripti | ei skeemaa | lista |
|-----------|----------|---------|------------|-------|
| 3. luokka | 0,5 | 1,3 | 1,1 | 0,6 |
| 4. luokka | 0,5 | 0,9 | 0,9 | 1,2 |
| 5. luokka | 0,4 | 0,9 | 1,4 | 0,3 |
| 6. luokka | 0,8 | 1,0 | 0,6 | 1,1 |

Vaikka numeroiden osuus muuten vaihtelee – erityisen paljon se vaihtelee listatason kirjoitelmissa – parhaissa eli kertomustason kirjoitelmissa se on suhteellisen vähäinen kaikkien muiden paitsi 6. luokan kertomuksissa. Numeroiden – ja yleensä numeraalien – käyttö näyttäisi siis lisääntyvän lasten varttuessa, eikä suhteellinen osuus saneista ole kovin selkeästi yhteydessä kirjoitustaitoon, vaikka vaikuttaisikin siltä, että taitavimmat kirjoittajat käyttävät niitä pääsääntöisesti vähiten.

Tytöt käyttävät numeroita poikiin verrattuna huomattavan vähän alemmilla luokilla, mutta ylemmillä luokilla tilanne tasoittuu. Suhteutettuna kirjoitelmien yhteissanemäärään numeroiden osuus 4.-luokkalaisten poikien kirjoitelmissa on 1,6 % ja tyttöjen kirjoitelmissa 0,6 %. 6.-luokkalaisilla pojilla numeroiden osuus on 1,2 % saneista ja tytöillä 0,7 % saneista.

Numeroiden käytöllä voi olla kirjoitustaidon lisäksi yhteys kielitaitoon. 4. luokan aineiston numeroista noin 40 % löytyi kielikylpykoululaisten teksteistä. Kuitenkin kielikylpykoululaisten sanemäärä kattaa koko 4. luokan saneista vain noin 23 %. Kielikylpykoululaisten kirjoitelmat ja samalla niiden sisältämät numerot on jätetty pois laskelmista, joissa on tarkasteltu eritasoisten aineiden numeromääriä, sillä heidän aineidensa tasoja ei ole luokiteltu. Numeroiden osuus 4. luokan kielikylpykoululaisilla on kuitenkin huomattavan iso, noin 1,6 % saneista. Toisaalta 3. luokan kielikylpykoululaisilla numeroiden osuus saneista ei ole merkittävä. Lisäselvitykset olisivatkin tarpeen, sillä kiistattomasti numeroiden käytön ja kielitaidon yhteyttä ei tässä voi osoittaa.

3.3 Yksilöllinen vaihtelu ja tasoerot

3.3.1 Sisältösanaluokkien ja pronomininien osuudet

Tämän luvun laskelmat on tehty aineistosta, josta 4.- ja 6.-luokkalaisten otannasta kielikylpyaineisto ja Raahen aineisto on poistettu aineistojen yhdenmukaistamiseksi. Karsinta johtuu osin siitä, että kielikylpy- ja Raahen aineiston kirjoitelmien tasoja ei ole määritetty. 2.-luokan aineistoon Raahen aineisto on jätetty, sillä alkuperäisen Koulukorpuksen, johon kirjoitelmien tasot on määritetty, osuus on 2. luokan aineistossa hyvin pieni. Tästä syystä 2.-luokan aineistosta ei myöskään tehdä kirjoitelmien tasoihin perustuvia laskelmia.

Ensin tarkastellaan sisältösanaluokkien eli substantiivien, verbien, adjektiivien ja adverbien, sekä pronomininien osuuksia 2.-, 4.- ja 6.-luokkalaisten teksteissä. Sen jälkeen tarkastelun kohteeksi otetaan substantiivien ja verbien suhde kirjoitelmissa.

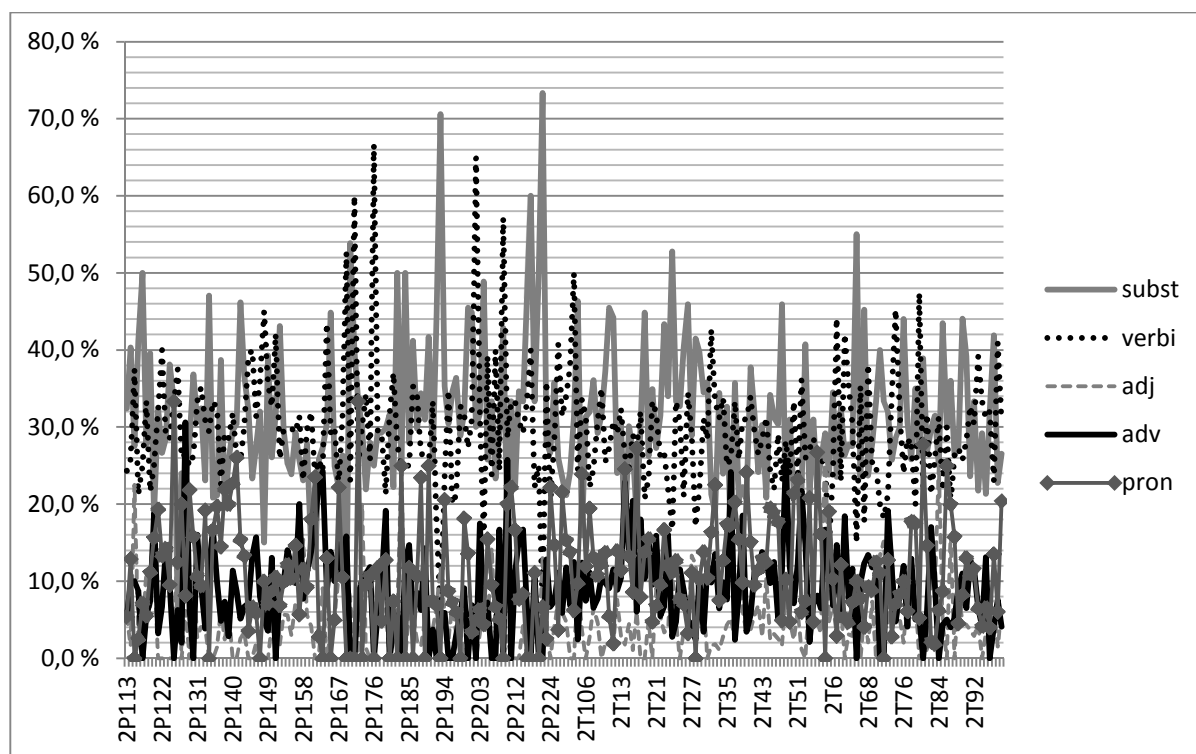
Vertailu eri luokkien välillä (taulukko 16) osoittaa, että substantiivien suhteellinen osuus kaikista sanoista pysyy samana, kun taas verbien osuus hieman pienenee varttuneemmilla kirjoittajilla. Adjektiivien ja adverbien osuus kasvaa ja pronomininien osuus hieman pienenee. Kaiken kaikkiaan näihin sanaluokkiin kuuluvien lekseemien osuudet tasaantuvat ajan myötä niin, että koko luokan tarkastelussa keskihajonta pienenee.

Tyttöjen väliset erot sanaluokkien osuuksissa ovat joka luokalla pienemmät kuin pojilla. Tyttöillä pronomininien ja adjektiivien osuudet pienenevät huomattavasti ylemmillä luokilla, adverbien käyttö sen sijaan lisääntyy. Pojilla verbien osuus pienenee jonkin verran, kun taas adverbien ja adjektiivien osuus kasvaa. Pronominien osuus on suurimmillaan 4. luokalla ja pienimmillään 6. luokalla.

Taulukko 16. Substantiivien, verbien, adjektiivien, adverbien ja pronomien osuuksien keskiarvot, mediaanit ja keskihajonnat 2., 4. ja 6. luokan oppilailla.

| | | 2. luokka | | | 4. luokka | | | 6. luokka | | |
|--------|--------------|-----------|-------|-------------|-----------|-------|-------------|-----------|-------|-------------|
| | | tytöt | pojat | kaikki | tytöt | pojat | kaikki | tytöt | pojat | kaikki |
| subst. | keskiarvo | 30,3 | 30,0 | 30,2 | 29,8 | 30,2 | 30,3 | 30,3 | 30,1 | 30,2 |
| | mediaani | 30,8 | 32,1 | 31,4 | 29,3 | 29,7 | 29,4 | 30,6 | 30,6 | 30,6 |
| | keskihajonta | 8,1 | 11,0 | 9,6 | 6,3 | 6,9 | 6,6 | 5,4 | 7,5 | 6,5 |
| verbi | keskiarvo | 28,4 | 29,6 | 29,2 | 29,0 | 28,2 | 28,6 | 28,3 | 28,4 | 28,3 |
| | mediaani | 28,8 | 30,5 | 29,7 | 28,6 | 27,8 | 28,4 | 28,4 | 28,3 | 28,3 |
| | keskihajonta | 6,6 | 9,2 | 8,0 | 5,2 | 5,5 | 5,4 | 4,4 | 5,2 | 4,8 |
| adj | keskiarvo | 9,3 | 3,4 | 3,7 | 6,3 | 4,8 | 5,6 | 6,1 | 5,4 | 5,8 |
| | mediaani | 9,4 | 4,5 | 4,8 | 5,6 | 4,4 | 5,4 | 6,4 | 5,7 | 6,0 |
| | keskihajonta | 5,3 | 5,1 | 4,9 | 3,7 | 4,0 | 3,9 | 3,4 | 3,7 | 3,6 |
| adv | keskiarvo | 9,3 | 7,7 | 8,7 | 10,8 | 10,3 | 10,6 | 11,8 | 10,8 | 11,4 |
| | mediaani | 9,4 | 8,4 | 8,9 | 10,5 | 10,8 | 10,7 | 12,0 | 11,2 | 11,6 |
| | keskihajonta | 5,3 | 6,5 | 5,9 | 4,0 | 5,6 | 4,9 | 3,7 | 4,6 | 4,2 |
| pron. | keskiarvo | 11,1 | 9,5 | 10,4 | 9,9 | 11,0 | 10,4 | 8,7 | 8,9 | 8,8 |
| | mediaani | 11,7 | 10,2 | 11,0 | 9,3 | 11,3 | 9,8 | 9,2 | 9,5 | 9,3 |
| | keskihajonta | 6,4 | 8,0 | 7,3 | 4,9 | 6,3 | 5,6 | 3,8 | 4,7 | 4,2 |

Seuraavaksi tarkastellaan sitä, kuinka paljon substantiivien, verbien, adjektiivien, adverbien ja pronomien osuudet kirjoitelmien saneista vaihtelevat yksilötasolla vuosiluokilla 2, 4 ja 6 (kuviot 19, 20 ja 21). Poikien kirjoitelmat sijaitsevat kuvioissa vasemmalla ja tyttöjen oikealla puolella, mikä auttaa hahmottamaan vaihtelua tyttöjen ja poikien välillä.



Kuvio 19. Substantiivien, verbien, adjektiivien, adverbien ja pronomien osuudet 2. luokan oppilailla.

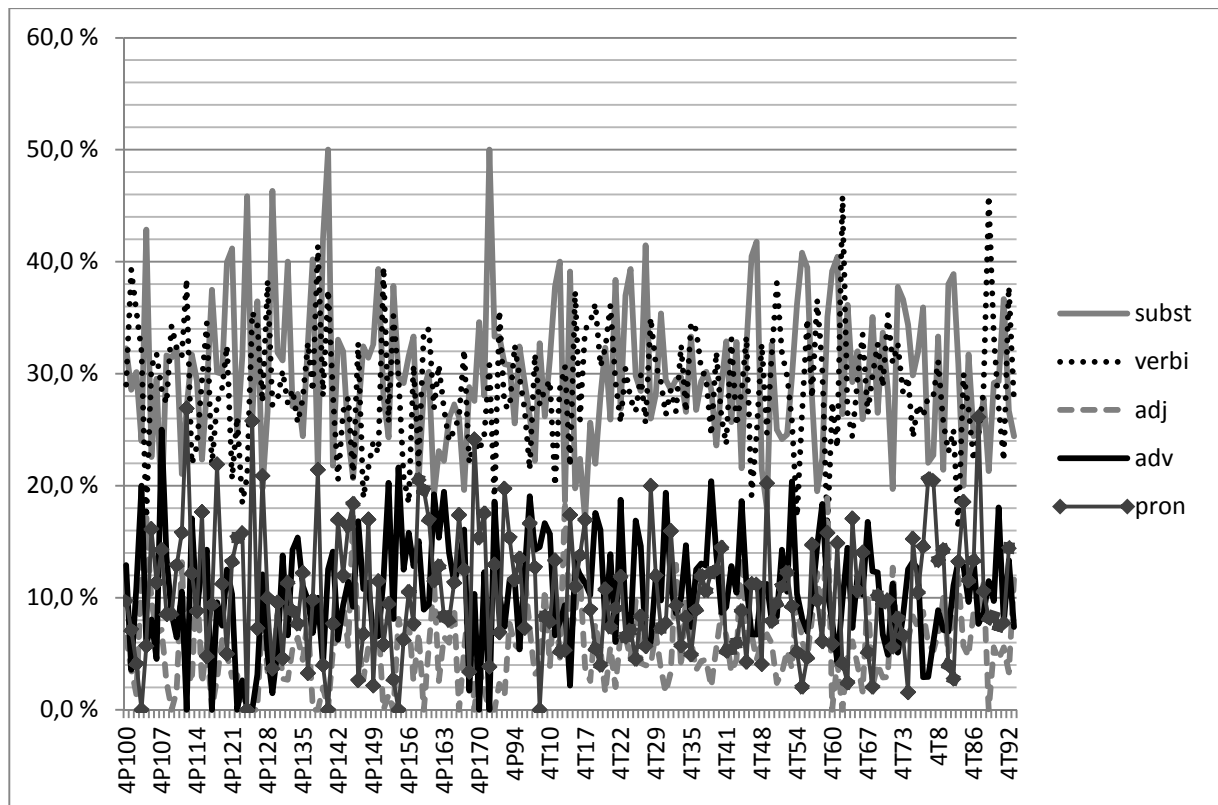
Yksilöllinen variaatio eri sanaluokkien osuuksissa on 2. luokan kirjoitelmissa huikeaa (kuvio 19). Substantiivien osuudet vaihtelevat välillä 5,3–73,3 %, verbien 5,9–66,7 %, adjektiivien 0–29,2 %, adverbien 0–30,6 % ja pronomien 0–33,3 %. Kirjoitelmia, joissa ei ole lainkaan adjektiiveja, on yhteensä 19. Niistä 16 on poikien ja loput tyttöjen. Adverbit puuttuvat täysin seitsemästä poikien kirjoitelmasta ja pronominit viidestä poikien kirjoitelmasta. Tuloksia tarkasteltaessa on toki otettava huomioon, että kirjoitelmien pituuskin vaihtelee suuresti (taulukko 1 s. 12). Neljän saneen mittaisessa kirjoitelmassa yksi sane edustaa 25 % kaikista saneista. Esimerkiksi eräässä viiden saneen aineessa substantiivien osuus on 60 % ja verbien 40 %. Muihin sanaluokkiin kuuluvia lekseimejä ei ole käytetty.

Aineessa (esimerkki 35), jossa substantiivien osuus on suurimmillaan (73,3 %), on 15 sanetta yhdyssanavirheiden poistamisen jälkeen. Vertailukohtana eräässä toisessa 15 saneen aineessa (esimerkki 36) substantiivien osuus on vain 26,7 %.

35) liikuntuu kiva koulu ruokailu televisiota kuvista tekniikkaa kässää evätensyönti kivakoulua näitä asioita tekisin (taajama 2P222)

36) Olipa KeRRAn mato ja SiTä yRiTteTiin syä mutta ei oLuThäTää se KuunteLi LOORDin Levyä (taajama 2P122)

Aivan yksi yhteen eivät siis kirjoitelmien pituus ja sanaluokkien osuudet mene. Näiden kirjoitelmien tasoa ei ole arvioitu, mutta esimerkin 35 kirjoitelma on listamainen, kun taas esimerkin 36 kirjoitelmassa on selvästi kerronnallisempi ote. Tosin 4. ja 6. -luokkien kirjoitelmien tarkastelun perusteella sanaluokkien osuudet eivät näytä olevan suoraan yhteydessä myöskään kirjoitelman tasoon.



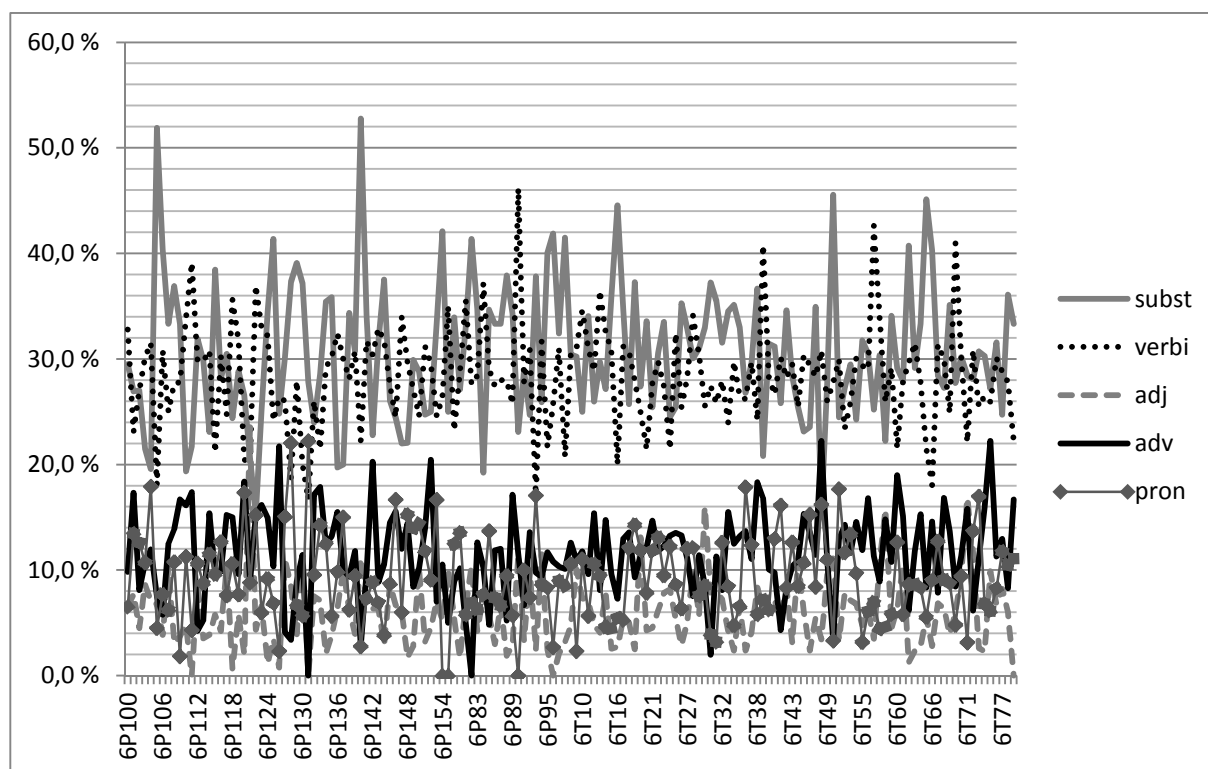
Kuvio 20. Substantiivien, verbien, adjektiivien, adverbien ja pronomienien osuudet 4. luokan oppilailla.

Eri sanaluokkien osuudet vaihtelevat paljon 4.-luokkalaisten kirjoitelmissa, vaikka 2. luokan kirjoitelmien suuruista vaihtelu ei enää olekaan (kuvio 20). Substantiivien osuudet vaihtelevat välillä 17,0–50,0 %, verbien 15,8–45,9 %, adjektiivien 0–19,3 %, adverbien 0–25,0 % ja pronomienien 0–26,9 %. Kirjoitelmia, joissa ei ole lainkaan adjektiiveja, on yhteensä 19. Niistä 16 on poikien ja loput tyttöjen. Adverbit puuttuvat täysin seitsemästä poikien kirjoitelmasta ja pronominit viidestä poikien kirjoitelmasta. Erään oppilaan kirjoitelmassa (esimerkki 37) ei ole lainkaan adjektiiveja, adverbeja tai pronomineja. Sen sijaan hänen kirjoitelmansa vilisee numeraaleja.

37) Kun heräsin, söin aamupalalla 11/2 suklaalevyä. Lähdin kahden tunnin päästä lounaalle viiden tähden ravintolassa. Kello kahdeksan äiti voitti lotossa ja yhdeksältä lähdimme lomalle Bancokc. (kaupunki 4P124)

Keskihajonnan perusteella (taulukko 16 s. 63) 4.-luokkalaiset tytöt käyttävät kaikkia tässä tarkasteltuihin sanaluokkiin kuuluvia sanoja tasaisemmin kuin pojat. Keskimäärin tytöt käyttävät substantiiveja, verbejä ja adverbeja hieman enemmän ja adjektiiveja selvästi

enemmän kuin pojat. Pronominien suhteellinen osuus sen sijaan on suurempi poikien kuin tyttöjen kirjoitelmissa.



Kuvio 21. Substantiivien, verbien, adjektiivien, adverbien ja pronominien osuudet 6. luokan oppilailla.

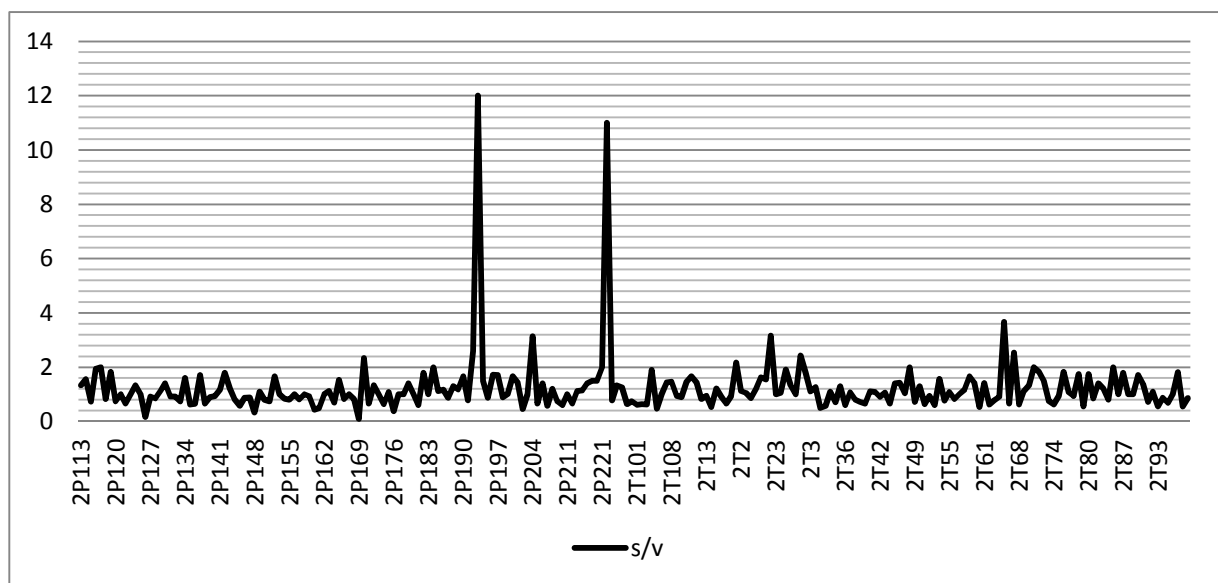
Eri sanaluokkien osuudet vaihtelevat paljon myös 6.-luokkalaisten kirjoitelmissa (kuvio 21). Toisaalta kuvion perusteella vaihtelu on 6. luokalla hieman pienempää kuin 4. luokalla, vaikka erot korkeampien huippujen ja syvimpien notkelmien välillä saattavatkin olla suurempia ainakin substantiivien ja adjektiivien osuuksissa. Substantiivien osuudet vaihtelevat välillä 15,2–52,8 %, verbien 16,7–46,2 %, adjektiivien 0–23,5 %, adverbien 0–22,2 % ja pronominien 0–22,2 %. Kirjoitelmia, joissa ei ole lainkaan adjektiiveja, on yhteensä neljä. Niistä kolme on poikien ja yksi tytön. Adverbejä ei esiinny kahdessa poikien kirjoitelmassa ja pronominit puuttuvat kolmesta poikien kirjoitelmasta.

Myös 6.-luokkalaiset tytöt käyttävät poikia tasaisemmin kaikkiin tässä tarkasteltuihin sanaluokkiin kuuluvia sanoja (taulukko 16). Substantiivien, verbien ja pronominien osuudet ovat tytöillä ja pojilla lähes samat. Adjektiivien ja adverbien osuus on tytöillä jonkin verran suurempi kuin pojilla.

Edellä oleva luokkien ja sukupuolten välinen tarkastelu osoittaa, että vaikka yksilöittäin eri sanaluokkien osuudet voivat vaihdella paljonkin, luokkien välillä kovin suuria eroja ei ole. Tytöt ovat kuitenkin jatkuvasti poikia tasaisempia kirjoittajia, mikä oli jo aiemmin tehtyjen tutkimusten valossa odotettavissa.

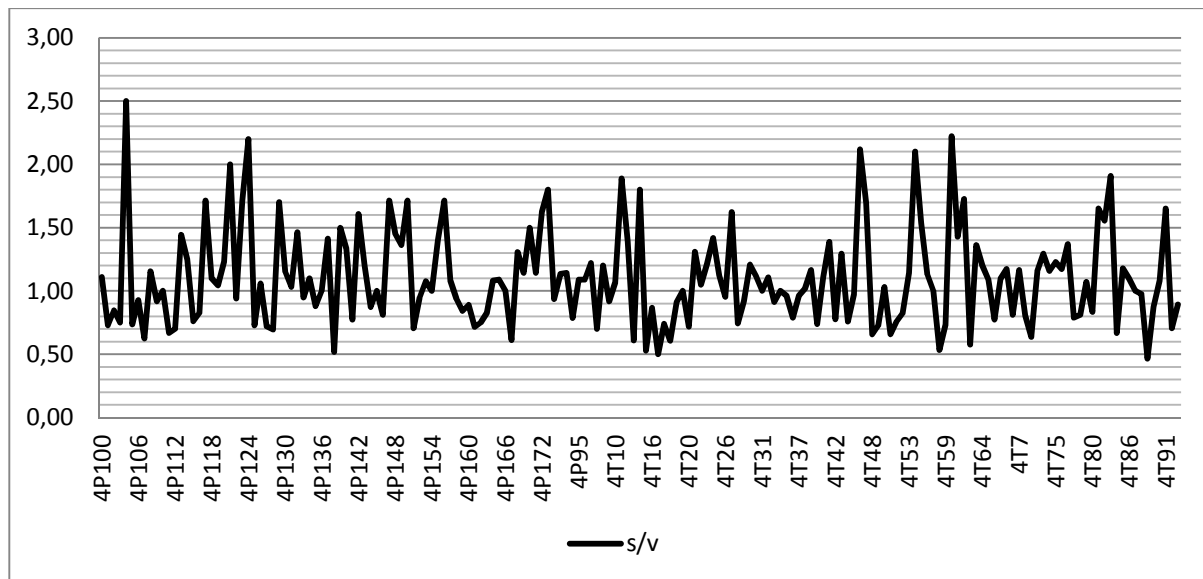
3.3.2 Substantiivi-verbi-suhde

Seuraavaksi tarkastellaan substantiivisaneiden määrän suhdetta verbisaneiden määrään 2.-, 4.- ja 6.-luokkalaisten kirjoitelmissa.



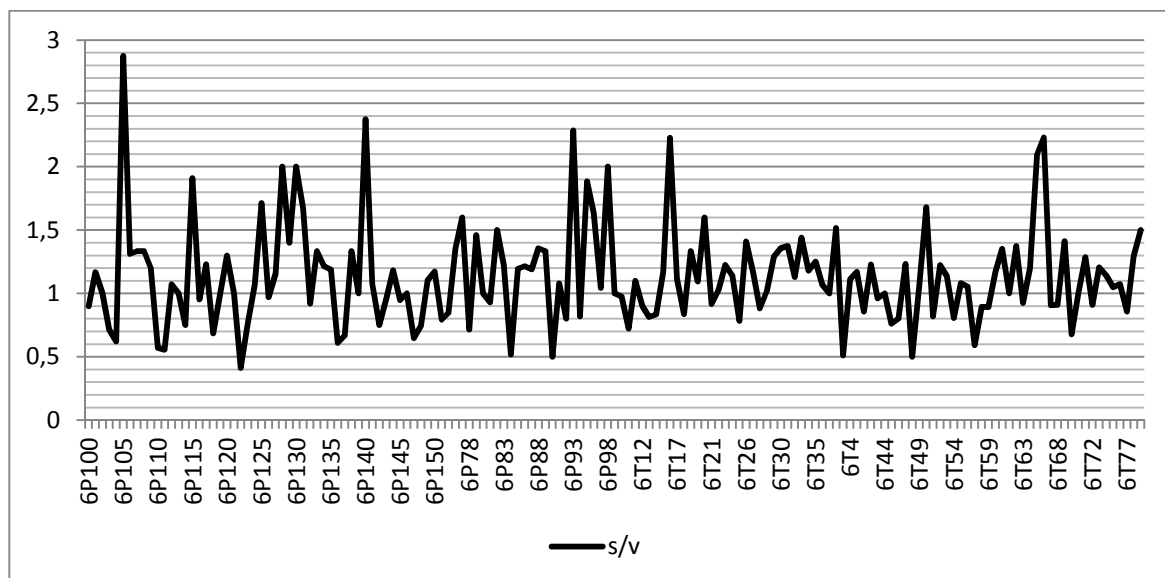
Kuvio 22. Substantiivien ja verbien suhde 2.-luokkalaisten kirjoitelmissa.

2.-luokkalaisten kirjoitelmissa substantiivien ja verbien suhde vaihtelee vielä paljon (kuvio 22). Poikien kirjoitelmissa suhde vaihtelee välillä 0,1–12 ja tyttöjen kirjoitelmissa välillä 0,47–3,67. Poikien kirjoitelmissa substantiivi-verbi-suhteen mediaani on 1, keskiarvo 1,29 ja keskihajonta 1,49. Tyttöjen kirjoitelmissa mediaani on 1,04, keskiarvo 1,16 ja keskihajonta 0,54. Tyttöjen kirjoitelmissa verbejä ja substantiiveja on siis tasaisemmin kuin poikien.



Kuvio 23. Substantiivien ja verbien suhde 4.-luokkalaisten kirjoitelmissa.

4.-luokkalaisilla substantiivi-verbi-suhde pysyy jonkin verran tasaisempuna kautta linjan (kuvio 23). Poikien kirjoitelmissa suhde vaihtelee välillä 0,52–2,5 ja tyttöjen kirjoitelmissa välillä 0,46–2,22. Poikien kirjoitelmissa substantiivi-verbi-suhteen mediaani on 1,08, keskiarvo 1,12 ja keskihajonta 0,38. Tyttöjen kirjoitelmissa mediaani on 1,03, keskiarvo 1,08 ja keskihajonta 0,38.



Kuvio 24. Substantiivien ja verbien suhde 6.-luokkalaisten kirjoitelmissa.

6.-luokkalaisilla suhdeluku on 2.-luokkalaisiin verrattuna huomattavan tasainen, mutta 4.-luokkalaisiin verrattuna ero poikien ja tyttöjen välillä kasvaa (kuvio 24). Myös hajonta

poikien aineistossa on suurempi kuin 4.-luokkalaisten poikien aineistossa, vaikka tyttöjen aineistoissa hajonta pienenee 6.-luokkalaisilla edelleen 4.-luokkalaisiin verrattuna. 6.-luokkalaisten poikien kirjoitelmissa substantiivi-verbi-suhde vaihtelee välillä 0,41–2,88 ja tyttöjen kirjoitelmissa välillä 0,5–2,23. Poikien kirjoitelmissa substantiivi-verbi-suhteen mediaani on 1,08, keskiarvo 1,16 ja keskihajonta 0,46. Tyttöjen kirjoitelmissa mediaani on 1,09, keskiarvo 1,12 ja keskihajonta 0,32.

Substantiivi-verbi-suhde näyttää pääsääntöisesti tasaantuvan lasten varttumisen myötä. 2.-luokkalaisten koko aineistossa substantiivi-verbi-suhteen mediaani on 1,00, keskiarvo 1,22 ja keskihajonta 1,11; 4.-luokkalaisten aineistossa mediaani on 1,05, keskiarvo 1,10 ja keskihajonta 0,38 ja 6.-luokkalaisten aineistossa mediaani on 1,08, keskiarvo 1,14 ja keskihajonta 0,40.

Miten substantiivi-verbi-suhde näkyy kirjoitelmassa? Mikäli suhdeluku on hyvin korkea, eli substantiiveja on paljon verbeihin verrattuna, kirjoitelma sisältää tavallisesti pitkiä substantiiviluetteloita (esimerkki 38).

38) Herään aamulla herätyskellolla kello seitsemältä. Pelaan tietokonetta 4 tuntia putkeen. Sitten syön lohitartaria kannullisen. Jälkiruoaksi syön Geishaa, Hornasisua, Pepeä, Tuplaa, Snickersiä, Marsia, Roopeankan kolikoita, Aarrearkkua, Marianne, Dumlea, irtokarkkeja, Omaria, Fazerin suklaalevyä, lakritsipötköä ja Tobleronea. Sitten matkustan Englannista Italian pelaamaan Uefan Campions Leaguen loppuottelun Villa Realia vastaan. -- (kaupunki 6P105; s/v 2,875)

Jos suhdeluku on alhainen, eli verbejä on paljon substantiiveihin verrattuna, käytetään kirjoitelmassa usein verbiketjuja ja substantiiveja korvataan pronomineilla (esimerkki 39):

39) Kun herään, leijun ilmassa. minulla ei ole kiire minnekään. Alan lukea kirjaa. Sitten yhtäkkiä pääsen kirjan maailmaan. Tapaan ystävällisiä leijonia ja osaan lentää. Saan ratsastaa leijonalla enkä tipu sen selästä. uin merien yli pelkäämättä koska meressä on vain delfiinejä. Syön ja juon metsäneläinten kanssa. Kohta he alkavat puhua, laulaa ja tanssia. He pyytävät minua tanssimaan ja suostun. Mutta kuinka Osasin tanssia? (maalais 4T89; s/v 0,46)

2.-luokkalaisten kirjoitelmissa alhainen suhdeluku voi johtua verbiuetteloista (esimerkki 40), jotka ylempien luokkien oppilaille enää harvoin dominoivat yhtä vahvasti kategorioiden välisiä suhteita.

40) Kun koulu alkaa taas on kivaa laske lukee kirjottaa ja oppia. Kun taas käymme uimassa niin on kivaa uida.

(kaupunki 2P169; s/v 0,1)

Tasoluokitusten kannalta substantiivi-verbi-suhteella ei näyttäisi olevan kovinkaan paljon merkitystä, sillä edellä näytetyt esimerkit osoittavat, että suhdelukujen ääripäissä voi esiintyä sekä varsin korkeatasoisia että myös heikkoja kirjoitelmia. Seuraavaan taulukkoon (taulukko 17) on koottu tiedot siitä, kuinka monet 2., 4. ja 6. luokan eritasoisista kirjoitelmista ovat substantiivi-verbi-suhteeltaan alle tai yli luokan mediaanin. Näyttää siltä, että listoissa kirjoitelmien substantiivi-verbisuhde sijoittuu epätasaisimmin mediaanin molemmin puolin vaihdellen yli 20 prosenttiyksikköä. Ainoastaan 4.-luokkalaisten skripteissä vaihtelu on myös tavallista suurempi.

Taulukko 17. Eritasoisien kirjoitelmien lukumäärät substantiivi-verbi-suhteen mukaan.

| | 2. luokka | | 4. luokka | | 6. luokka | |
|------------------|------------|------------|------------|------------|------------|------------|
| s/v | < mediaani | ≥ mediaani | < mediaani | ≥ mediaani | < mediaani | ≥ mediaani |
| kertomus (kpl) | 13 (45) | 16 (55) | 20 (50) | 20 (50) | 24 (48) | 26 (52) |
| skripti (kpl) | 28 (50) | 28 (50) | 46 (60) | 31 (40) | 31 (54) | 26 (46) |
| ei skeemaa (kpl) | 13 (42) | 18 (58) | 9 (47) | 10 (53) | 14 (58) | 10 (42) |
| lista (kpl) | 23 (39) | 36 (61) | 14 (33) | 28 (67) | 7 (32) | 15 (68) |

Eri sanaluokkien osuudet ja substantiivi-verbi-suhde voivat vaihdella huomattavasti samanikäistenkin oppilaiden teksteissä. Vaikka vaihtelu on suurta vielä 6. luokallakin, yleisesti ottaen oppilaiden väliset erot tasaantuvat ajan mittaan tasoluokitusten näkökulmasta.

Loppujen lopuksi substantiivi-verbi-suhteesta ei voi päätellä kovinkaan paljon, ja sitä pitäisi tutkia monimuuttuja-analyysillä, jotta mahdolliset riippuvuussuhteet tulevat esiin. Verbien määrä kertoo kirjoitelmasta substantiivien määrää enemmän, sillä alakouluikäisillä ei juuri vielä esiinny kehittyneemmille kirjoittajille tyypillistä lauseiden tiivistämistä.

3.4 Sanaston rikkaus

Tämän luvun laskelmiin on käytetty supistettua aineistoa, josta puuttuu kielikylpyaineisto ja 4. ja 6. luokan aineistosta myös Raahen aineisto, sillä niiden kirjoitelmien tasoja ei ole arvioitu. Myös numerot on jätetty pois, sillä niiden jättäminen antaisi sanaston monipuolisuudesta tunnuslukujen valossa turhan rikkaan kuvan. Käytetyn aineiston kokonaissanemäärä ynnä muut tarpeelliset tiedot ovat taulukossa 18.

Taulukko 18. Tietoja rikkauslukulaskelmiin käytetystä aineistosta.

| | 2. luokka | | 4. luokka | | 6. luokka | |
|------------------------|-----------|----------|-----------|----------|-----------|----------|
| | tytöt | pojat | tytöt | pojat | tytöt | pojat |
| oppilaiden määrä | 115 | 109 | 94 | 83 | 76 | 77 |
| saneita yhteensä | 6454 | 3960 | 8638 | 5341 | 10552 | 6553 |
| lekseemejä yhteensä | 1320 | 984 | 1754 | 1208 | 2123 | 1540 |
| subst.saneet/lekseemit | 1950/735 | 1239/564 | 2596/923 | 1614/664 | 3230/1139 | 2046/882 |
| verbisaneet/lekseemit | 1870/243 | 1179/173 | 2523/385 | 1521/237 | 3039/444 | 1883/283 |
| adj.saneet/lekseemit | 298/99 | 166/65 | 538/165 | 259/98 | 648/194 | 349/118 |
| adv.saneet/lekseemit | 653/139 | 381/98 | 950/197 | 570/128 | 1283/268 | 750/186 |

Tarkasteltaessa kirjoitelmien sanastoa tasoluokitusten näkökulmasta, käytetään aineistoa, jonka tunnusluvut näkyvät taulukossa 19.

Taulukko 19. Tietoja tasoluokitusten rikkauslukulaskelmiin käytetystä aineistosta.

| | 4. luokka | | | | 6. luokka | | | |
|------------------|-----------|---------|------------|-------|-----------|---------|------------|-------|
| | kertomus | skripti | ei skeemaa | lista | kertomus | skripti | ei skeemaa | lista |
| oppilaiden määrä | 40 | 76 | 19 | 42 | 50 | 57 | 24 | 22 |
| tyttöjä | 34 | 41 | 3 | 16 | 36 | 27 | 11 | 2 |
| poikia | 6 | 35 | 16 | 26 | 14 | 30 | 13 | 20 |
| saneita yhteensä | 4726 | 6161 | 836 | 2256 | 7647 | 6300 | 2176 | 982 |

3.4.1 Yhden kerran esiintyvät sanat

Yhden kerran esiintyvien sanojen eli HL-sanojen (hapaks legomenon) voisi olettaa kertovan jotakin tekstin sanastollisesta rikkaudesta ja kirjoittajan laajasta sanavarastosta. Niemikorven (1991: 79) mukaan HL-sanojen osuus riippuu paljolti otoksen koosta. Hän viittaa tutkimukseen, jonka mukaan pienissä, alle 2 000 saneen otoksissa HL-sanojen osuus kaikista saneista voi nousta jopa 80 %:iin. Niemikorven (1991: 80) lähes 300 000 saneen korpuksessa HL-sanojen osuus oli noin 22 %.

Niemikorven tutkimustuloksiin verrattuna Koulukorpus-aineistossa esiintyy hyvin vähän HL-sanoja, samoin myös verrattuna 2. ja 3. -luokkien oppikirjojen sanastoon (Jaakola 2004). Vaikka 1.-luokkalaisten aineisto sisältää 2 700 sanetta, eli varsin vähän, HL-sanojen osuus on vain 17 % saneista. Tämä voi johtua esimerkiksi aineiston luonteesta: koostuuhan se samasta aiheesta kirjoitetuista kouluaineista. Muilla luokilla HL-sanojen osuus on vielä pienempi (taulukko 20). Koko aineistossa HL-sanoja on noin 4 030 kun numerot jätetään laskelman ulkopuolelle, eli koko aineistossa HL-sanojen osuus on vain noin 5 %. HL-sanojen pieni osuus koko aineistossa johtuu osittain siitä, että HL-sanojen määrä on yleensä sitä pienempi, mikä isompi aineisto on kyseessä. Koko aineistossa HL-sanojen osuus on kuitenkin harvinaisen suuri, mikä saattaa johtua ainekin aineiden aihepiirien yhtenäisyydestä, jonka vuoksi oppilaat käyttävät paljon samoja sanoja, runsaasta toistosta yhden oppilaan kirjoitelmassa sekä yleisesti siitä, että nuoret kirjoittajat turvautuvat hyvin pitkälti yhteiseen, hyvin hallittuun perussanastoon.

Taulukko 20. HL-sanojen osuudet luokittain.

| | 1. luokka | 2. luokka | 3. luokka | 4. luokka | 5. luokka | 6. luokka |
|--|-----------|------------|------------|------------|------------|------------|
| Kerran esiintyviä lekseemejä (kpl/saneita) | 447/2661 | 1007/10414 | 1193/10807 | 1271/13979 | 1297/15904 | 1587/17105 |
| % saneista | 17 | 10 | 11 | 9 | 8 | 9 |
| % lekseemeistä | 63 | 55 | 58 | 54 | 55 | 55 |
| Tytöt HL-sanat (kpl/saneita) | 282/1762 | 612/6454 | 587/6031 | 795/8638 | 768/9413 | 973/10552 |
| HL%S | 16 | 9 | 10 | 9 | 8 | 9 |
| HL-sanat/kirjoitelma | 5 | 5 | 8 | 8 | 9 | 13 |
| Pojat HL-sanat (kpl/saneita) | 165/899 | 395/3960 | 606/4776 | 476/5341 | 529/6491 | 614/6553 |
| HL%S | 18 | 10 | 13 | 9 | 8 | 9 |
| HL-sanat/kirjoitelma | 3 | 4 | 7 | 6 | 6 | 8 |

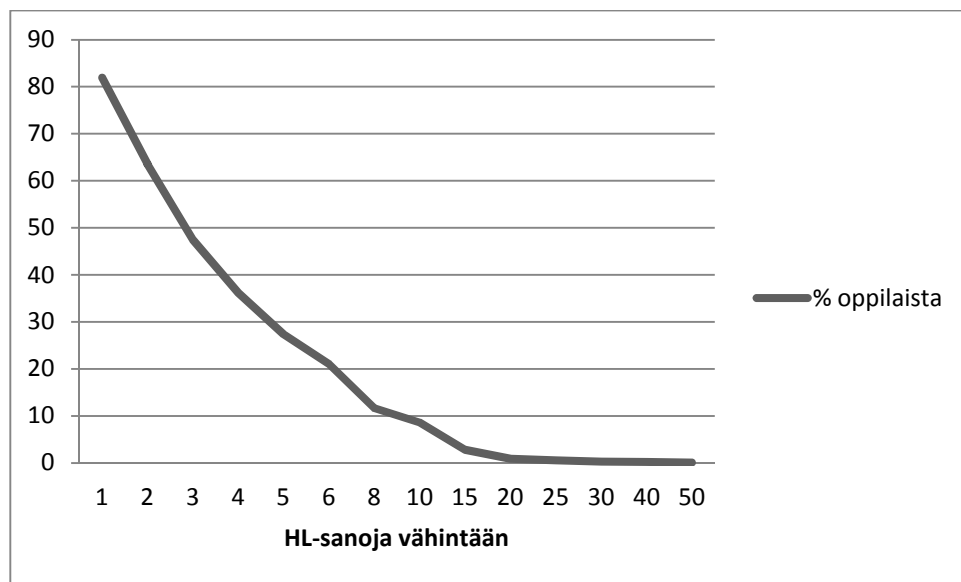
Vuosiluokkien välisessä vertailussa HL-sanojen osuuksissa ei näyttäisi tapahtuvan suuria muutoksia ensimmäisen luokan jälkeen, jolloin HL%S-arvo eli hapaks legomenon -sanojen suhde saneisiin pienenee huomattavasti (taulukko 20). Pientä vaihtelua kuitenkin esiintyy, silmiinpistävämpänä ehkä 3.-luokkalaisten poikien suuri HL%S-arvo, mitä voisi selittää kielikylypykirjoitelmilla. Myös tyttöjen ja poikien välinen ero kirjoitelmien keskimääräisessä HL-sanojen määrässä kasvaa selvästi 4.-luokalta alkaen.

Sanemääriin suhteutettuna 1.-3.-luokkalaisilla pojilla on hieman enemmän HL-sanoja kuin tyttöillä. On kuitenkin kyseenalaista, kuinka hyvin HL-sanojen määrä kuvaa sanaston rikkautta, sillä ongelmana on se, että jos vain yksi kirjoittaja on käyttänyt tiettyä

lekseemiä x , mutta hän on käyttänyt sitä kahdesti, tätä lekseemiä ei enää käsitellä HL-sanana. Onko kirjoitelman sanasto heti niukempi, mikäli jokin lekseemi esiintyy siinä useammin kuin kerran, vaikka kyseistä lekseemiä ei esiintyisi kenenkään toisen kirjoitelmassa?

Luokkien HL%S-arvon perusteella 1.-luokkien kirjoitelmien sanasto on monipuolisinta, mikä on järjenvastaista. Jaakolakin (2004: 58) toteaa, että sanemäärältään hyvin suppeissa aineistoissa HL-sanoja on suhteellisen paljon siksi, että yksittäisten sanojen toistoa on vähän. Mitä pidempi kirjoitelma on, sitä todennäköisemmin toistoa siis esiintyy, ja tyttöjen kirjoitelmat ovat keskimäärin poikien kirjoitelmia pidempiä. Esimerkiksi 2.-luokkalaisten tyttöjen kirjoitelmien keskipituus on 56 ja poikien 37 sanetta, 5.-luokkalaisten tyttöjen kirjoitelmien keskipituus taas 129 ja poikien 89 sanetta. Kirjoitelmien määriin suhteutettuna HL-sanoja esiintyy kuitenkin enemmän tyttöjen kuin poikien kirjoitelmissa ja tästä näkökulmasta HL-sanojen määrä myös lisääntyy, kun siirrytään ylemmille vuosiluokille.

HL-sanojen määrä yhden oppilaan tekstissä vaihtelee melko paljon. Koska numerot eivät ole HL-sanoina erityisen kiinnostavia, ne on poistettu aineistosta seuraavassa tarkastelussa (kuvio 25). Noin 82 %:lla oppilaista on kirjoitelmassaan vähintään yksi HL-sana, mutta vähintään kolme sanaa on enää vain vajaalla 50 %:lla ja vähintään kymmenen sanaa alle 10 %:lla. Yli 50 HL-sanaa on vain yhden oppilaan tekstissä.



Kuvio 25. HL-sanojen vähimmäismäärä kirjoitelmassa x %:lla oppilaista.

Seuraavaksi tarkastellaan HL-lekseemien määrää suhteessa sanemäärään ja kirjoitelmien määrään rajatussa aineistossa (taulukko 18 s. 71). Koska luvun 3.4.2 indeksilaskelmissa 2., 4.

ja 6. luokkien tyttöjä ja poikia käsitellään omina ryhminään, määritetään HL-arvot myös erikseen tyttöjen ja poikien kirjoitelmista. Tämä tarkoittaa sitä, että tyttöjen ja poikien aineistoissa voi HL-sanoina esiintyä sellaisia sanoja, jotka eivät ole HL-sanoja koko luokan tai koko aineiston tarkastelussa.

Taulukko 21. HL-sanojen määrä suhteutettuna kirjoitelmien sanemääriin (TTR-arvo).

| | 2. luokka | 4. luokka | 6. luokka |
|-----------|-----------|-----------|-----------|
| Tytöt (%) | 11 | 11 | 11 |
| Pojat (%) | 15 | 13 | 14 |

Kun tyttöjen ja poikien kirjoitelmia tarkastellaan omina aineistoinaan, HL-sanojen määrä ryhmittäin ja samalla myös TTR-arvo ovat hieman suurempia kuin luokittaisessa vertailussa (taulukko 21). Samansuuntaisia tulokset kuitenkin edelleen ovat: Tyttöjen arvot ovat joka luokalla yhtä suuret, poikien arvot eroavat toisistaan hivenen ja ovat joka luokalla tyttöjen arvoja suurempia.

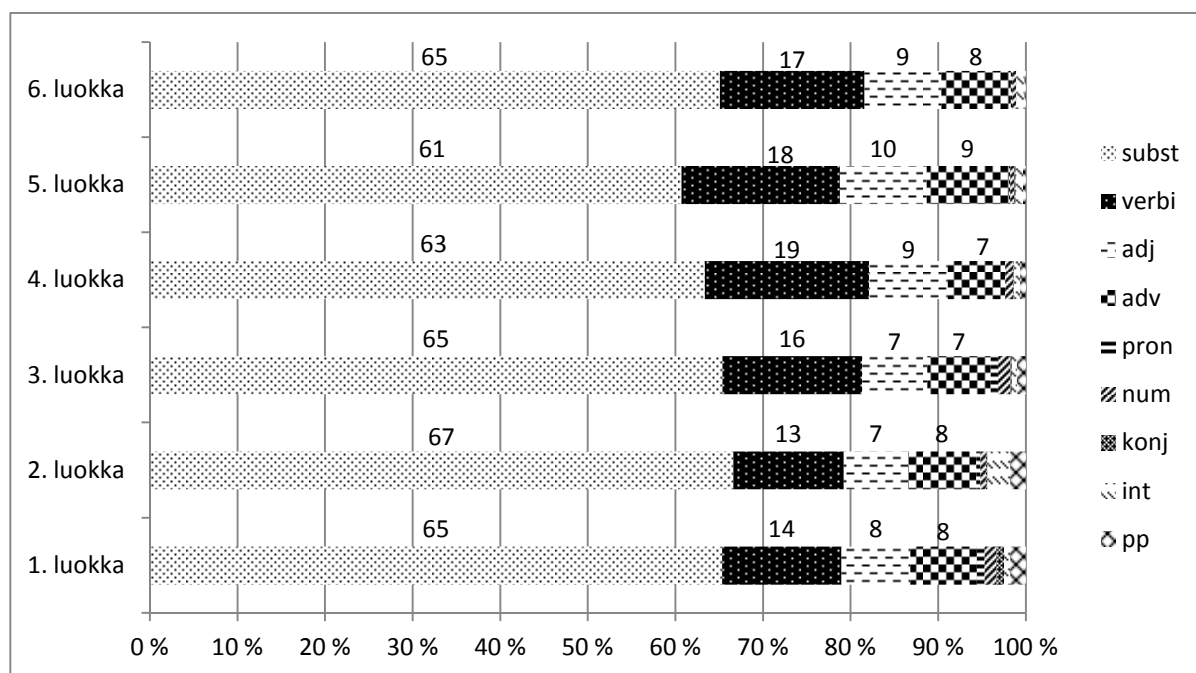
Jos HL-sanojen määrä sen sijaan suhteutetaan kirjoitelmien määrään, tulos on täysin erilainen (taulukko 22). Tyttöjen kirjoitelmissa on keskimäärin enemmän HL-sanoja kuin poikien kirjoitelmissa. HL-sanojen määrä myös kasvaa selvästi, kun siirrytään ylemmille luokka-asteille. Voidaan siis todeta, että HL-sanojen määrä kasvaa kirjoitelmien sanemäärän kasvaessa, mutta ei suhteessa saneisiin.

Taulukko 22. HL-sanojen määrä yhdessä kirjoitelmassa keskimäärin.

| | 2. luokka | 4. luokka | 6. luokka |
|-------------|-----------|-----------|-----------|
| Tytöt (kpl) | 6 | 10 | 15 |
| Pojat (kpl) | 5 | 8 | 12 |

HL-sanojen suhde sanemääriin ei ole ainoa, eikä varsinkaan luotettavimmalta vaikuttava menetelmä sanaston rikkauten arvioimiseksi ainakaan tässä aineistossa. Sanaston rikkautta eli TTR-arvoa, jossa keskimääräisiä sanaluokkaesiintymiä suhteutetaan sanemääriin, on käyttänyt esimerkiksi Saarela (1991). Myös hänen tulostensa perusteella pojat käyttäisivät suhteellisesti runsaammin erilaisia sanoja kuin tytöt. Saarela (1991: 24) mainitsee TTR-arvon heikkoudeksi sen, että se ei ota huomioon aineiston hajontaa ja luotettavuuden vähenemistä, kun aineiston sanemäärä on alle 10 000 sanetta. Sanaston rikkautta tutkittaessa ei ole siis lainkaan merkityksetöntä, mihin tuloksia suhteutetaan ja mitä menetelmiä käytetään. HL-sanojen määrän vertailua hienostuneempiakin laskennallisia menetelmiä sanaston

monipuolisuuden kuvaamiseen on olemassa. Niitä sovelletaan luvussa 3.4.2 Indeksit sanaston monipuolisuuden määrittelyssä.



Kuvio 26. Yhden kerran esiintyvien lekseemien sanaluokat luokittain.

Suurin osa HL-sanoista kuuluu substantiivien kategoriaan (kuvio 26). Seuraavaksi suurin kategoria on verbit ja sen jälkeen adjektiivit ja adverbit. Yhdessä nämä neljä suurinta luokkaa kattavat 2.-luokkalaisten HL-sanoista 95 %, 4.-luokkalaisten HL-sanoista 98 % ja 6.-luokkalaisten HL-sanoista jopa 99 %. 1.-luokan aineistosta HL-sanoja löytyy vähintään 1 % jokaisesta yhdeksästä kategoriasta, mutta jo 2.-luokalla HL-sanoja löytyy neljän yleisimmän kategorian lisäksi enää vain kolmesta kategoriasta. Nämä luvut kertovat siitä, että alimmilla luokilla eivät kaikki tavallisimmat pronominit tai konjunktioita ole vielä kaikkien käytössä.

Numeraalien osuus HL-sanoista on suurin 3. luokalla, interjektioiden 2. luokalla. Konjunktioiden käytön kehittymisestä kertoo se, että 6.-luokkalaisilla vain kerran käytettyjä konjunktioita ei enää esiinny.

Useat HL-sanat ovat joko propreja kuten henkilöiden, joukkueiden tai paikkojen nimiä, pitkiä yhdyssanoja, pisimpinä esimerkkeinä *adventtipiparkakkutalo* ja *ajokortinsaanti-ikä*, tai vierasperäisiä sanoja, kuten sitaatti- tai muita lainoja, mutta niiden lisäksi löytyy muun muassa myös lyhyehköjä sanoja, jotka erottaa paljon käytetyistä sanoista ainakin se, että ne eivät kuulu semanttisesti perustasolle, vaan ovat perustason alapuolisia sanoja, kuten lajinnimiä. Tällaisia sanoja ovat esimerkiksi *ahven*, *kantarelli*, *karitsa*, *kettuterrieri*, *lokki* ja

makrilli. Esimerkkinä lainasanoista ovat *cool*, *dipata*, *hengata*, *jockey* ja *teleportata*. Harrastuksiin liittyvää erikoissanastoa ovat esimerkiksi *connemara*, *EM-estekilpailu*, *ratsastuskypärä* ja *ruunikonkimo* (hevoset), *uistin*, *virvelöidä* (kalastus) sekä *formula*, *futisstadion*, *kroolata*, *rannelaukaus* ja *sarjajumbojoukkue* (urheilu). Toki harrastussanasto sisältää usein paljon lainasanoja. Muita HL-sanoja ovat esimerkiksi ”lipsahdukset”, kuten *hengenvartija* (*henkivartija*), nuorisokielen ilmaukset ja muut hyvään yleiskieleen kuulumattomat sanat, kuten *biitsi*, *bilettää*, *hokata* ja *hottis* tai jopa alatyyliset sanat kuten *kusella* ja *pallit* sekä interjektiot, kuten *aaargh*, *aijai*, *jep* ja *ööö*.

HL-verbien tyypillisiä ominaisuuksia ovat johtimellisuus, esimerkiksi *ajattaa*, *ajautua*, *ajelehtia*, deskriptiivisyys tai ekspressiivisyys, esimerkiksi *hihhuloida*, *kipittää*, *käppäillä*, *tallustella*, *tupruta* ja *älämölistä*, ja emotionaalisuus kuten *hurmaantua*, *ilostua*, *ikävoidä*, *säikähtää* ja *tympiä*. Toki HL-sanojen joukkoon mahtuu varsin tavalliseltakin tuntuvia, lyhyehköjä sanoja, esimerkiksi *maksu* ja *reitti*, mikä voi johtua ainakin kirjoitelmien aiheen vaikutuksesta sanastoon, kirjoittajien iästä sekä aineiston koosta.

HL-sanojen joukossa on myös aineen otsikon ja kirjoittajien iän huomioon ottaen hieman odottamattomia sanoja, esimerkiksi *hanuri*, *hepatiitti*, *hautajaiset*, *pervo*, *siideri*, *surmata* ja *sydänkohtaus*, joiden käyttöyhteys näkyy seuraavissa lauseissa (esimerkit 41–47):

41) Oot sääki aika hanurista. (5BP9)

42) Joudumme ottamaan silti rokotteen joka estää hepatiitin. (kaupunki 6P108)

43) Otin singon ja konekiväärin ja pian oli hautajaiset. (kaupunki 3P122)

44) Joo hei nyt kaikki sit joku pervo ilme. (5BT6)

45) Ja oon saanu jopa alkoholitonta siiderii. (kaupunki 1T23)

46) Olla prinssi ja surmata lohikäärme. (kaupunki 3P 115)

47) Aamu alkoi mukavasti porealtaassa kun suuri limainen sammakko hyppäsi altaasen olin vähällä saada sydänkohtauksen (kielikylpy 4P7)

Esimerkkilauseet antavat hyvin kuvaa siitä, kuinka eri tavoin unelmien päivästä voidaan kirjoittaa. Vaikka annettu aihe vaikuttaakin paljon siihen, minkälaisena alakouluikäisten käyttämä sanasto tässä tutkimuksessa esittäytyy, antaa aihe myös mahdollisuuden mielikuvituksen käyttöön, mikä tuo yllättävilläkin tavoilla esiin alakouluikäisten sanaston rikkautta. Osittain HL-sanastoa rikastuttavat myös aineet, joiden sisältö ei vastaa otsikkoa.

HL-sanojen määrä ei kuitenkaan lopulta näytä paljonkaan korreloivan kirjoitustaidon kanssa ja niiden tarkastelu panee myös miettimään, kuinka paljon oppilaat miettivät tekstin lukijaa tai sitä, minkälainen kieli koulukontekstiin soveltuu. Kuinka uusi ilmiö on, että alakoululaisen kirjoitelmasta on löydettävissä edellä mainittuja alatyylisiä sanoja ja ilmauksia?

Lopuksi tarkastellaan vielä 4. ja 6. luokan HL-sanojen sijoittumista eri tasoluokkien kirjoitelmiin (taulukko 23). Laskelmiin on käytetty rajattua aineistoa (taulukko 19 s. 71).

Taulukko 23. HL-sanojen osuudet eritasoisissa kirjoitelmissa.

| | 4. luokka | 6. luokka |
|------------|-----------|-----------|
| kertomus | 11 | 10 |
| skripti | 8 | 9 |
| ei skeemaa | 10 | 7 |
| lista | 8 | 8 |

Kirjoitelman taso ei juuri näy HL-sanojen osuudessa muuten kuin siten, että kertomus-tasoisissa kirjoitelmissa HL-sanojen osuus on suurin, kun verrataan koko luokan HL-sanojen osuuksia eritasoisissa kirjoitelmissa. Muiden tasojen kohdalla HL-sanojen osuudet eivät vähene lineaarisesti kirjoitelmien tason heiketessä. Tähän voi vaikuttaa se, että 4. luokalla ei skeemaa -tasoisien kirjoitelmien ja 6. luokalla lista-tasoisien kirjoitelmien yhteissanemäärä on ollut huomattavasti pienempi kuin muun tasoisissa kirjoitelmissa. 6.-luokkalaisten lista-tasoiset kirjoitelmat ovat myös hyvin erilaisia alempien luokkien lista-tasoiisiin kirjoitelmiin verrattuna, jolloin niiden heikompi taso ei välttämättä näy mitenkään sanaston niukkuutena. Ei skeemaa -tasoiset kirjoitelmat taas ovat yleensäkin hyvin heterogeeninen joukko.

3.4.2 Indeksit sanaston monipuolisuuden määrittelyssä

Kirjoitelmien sanastoa on kuvattu jo lekseemitasolla tarkastelemalla yleisimpien lekseemejä ja niiden osuuksia. Myös sanaston monipuolisuuden ja sen muutosten tarkastelu luokka-asteittain sekä tyttöjen että poikien välillä on yksi tapa kuvata kielenkehitystä. Edellisessä luvussa tarkasteltiin HL-sanoja ja HL-sanojen suhdetta saneiden määrään. Kaikkein yksinkertaisin sanaston rikkauden määrittelyyn käytettävä mittari kuitenkin on TTR-arvo (=type-token-ratio), joka tarkoittaa lekseemien ja saneiden suhdetta. Sanaston monipuolisuuden määrittämiseen on TTR-arvon lisäksi useita muita laskennallisia indeksejä.

Särkkä (1987) on tutkinut sanaston runsauden mittareiden toimivuutta kaunokirjallisten teosten sanaston rikkauden määrittelyssä. Hän (1987: 131) jakaa sanavaraston runsauden mittarit kahteen ryhmään: sanojen ja saneiden suhteisiin perustuviin kertoimiin, joita ovat esimerkiksi TTR-kerroin ja Pierre Guiraud’n kehittämä R-indeksi (R=richesse) sekä hajontaan perustuviin indekseihin, joihin kuuluu esimerkiksi Herdanin V_m -indeksi. Omassa tutkimuksessaan Särkkä on käyttänyt TTR-arvoa, V_m -indeksiä ja HL-sanoja.

Saarela (1997) käyttää sanaston monipuolisuuden määrittelyn mittarina Shannon-Weaverin diversiteetti-indeksiä eli Shannonin entropiaa. Hän pitää sen vahvuutena sitä, että se ottaa huomioon sanaston hajonnan, korjaa suuritaajuisien sanojen aiheuttaman vinouman ja sopii myös pieniin aineistoihin, mitkä ovat hyviä perusteita sen käyttämiseen Koulukorpus-aineiston sanaston tutkimuksissa. Käyttämällä sanaston rikkauden määrittelyyn Shannonin entropiaa tuloksia voidaan verrata Saarelan saamiin tuloksiin.

Shannonin entropia voidaan määrittää seuraavalla kaavalla:

$$H' = - \sum_{i=1}^k p_i \ln p_i$$

Kaavassa H' merkitsee entropiaa, p_i merkitsee eri lekseemien suhteellista frekvenssiä ja k jokaista lekseemiä (ts. lekseemien määrää) (esimerkki kaavan soveltamisesta ks. esim. Saarela 1997:53). Entropian lisäksi voidaan seuraavan kaavan avulla määrittää entropian maksimiarvo H'_{max} :

$$H'_{max} = \ln k.$$

Jakauman tasaisuuden mittarina käytetään suhdelukua J' , joka määritetään seuraavasti:

$$J' = \frac{H'}{H'_{max}}$$

J' :n suurin arvo on 1, eli kyse on siitä, kuinka lähellä entropia on suurinta mahdollisinta arvoaan. Mitä lähempänä lukua 1 J' on, sitä tasaisemmin aineistossa eri lekseemejä käytetään (kaikki em. kaavat: ks. esim. Saarela 1997: 52).

Edellä mainittujen lisäksi muita sanaston rikkauden määrittelyyn käytettäviä indeksejä ovat esimerkiksi Simsonin indeksi ja entropia, joita on käyttänyt esimerkiksi Niemikorpi (1991) sekä Guiraud’n rikkausindeksi, TTR_{Carr} , TTR_{log} , Brunet’n W-indeksi, Honorén R-indeksi ja Herdanin V_m -indeksi, joita puolestaan on käyttänyt ainakin Jaakola (2004). Näistä indekseistä

tähän tutkimukseen valitaan Herdanin V_m -indeksi, sillä Särkän (1987: 134) mukaan tyyliältään yhtenäistä tekstilajia tarkasteltaessa se toimii luotettavasti otoskoosta riippumatta. Valintaan on kaksi syytä: Ensinnäkin 2.-luokkalaisten poikien aineistossa sanemäärä on reilusti alle 5000, eikä sanemäärä toki kovin suuri ole muidenkaan ryhmien aineistoissa. Lisäksi kouluaineiden pitäisi ainakin periaatteessa edustaa yhtenäistä tyyliä. Herdanin V_m -indeksi saadaan seuraavalla kaavalla:

$$V_m = \frac{S/X}{\sqrt{N}},$$

missä S on lekseemien frekvenssin keskiarvon keskivirhe, X lekseemien keskimääräinen frekvenssi ja N lekseemien määrä.

TTR_{Carr} -indeksin väitetään Jaakolan (2004: 101) mukaan tasoittavan aineiston kokoerojen aiheuttamaa vääristymää. Samoin pitäisi toimia myös TTR_{log} -indeksin, jossa logaritmien olisi tarkoitus tasoittaa aineiston sanemääristä johtuvia eroja (mts.). Pikaisen kokeilun perusteella TTR_{Carr} näyttäisi todellakin toimivan, sillä se antaa ylemmille luokille suurempia arvoja kuin alemmille. TTR_{log} -indeksin antamat luvut sen sijaan näyttävät samansuuntaisilta kuin TTR -indeksin. TTR_{Carr} -indeksin valinta tähän tutkimukseen vaikuttaa siis TTR_{log} -indeksiä perustellummalta. TTR_{Carr} -indeksi lasketaan seuraavasti:

$$TTR_{Carr} = \frac{L}{\sqrt{2S}},$$

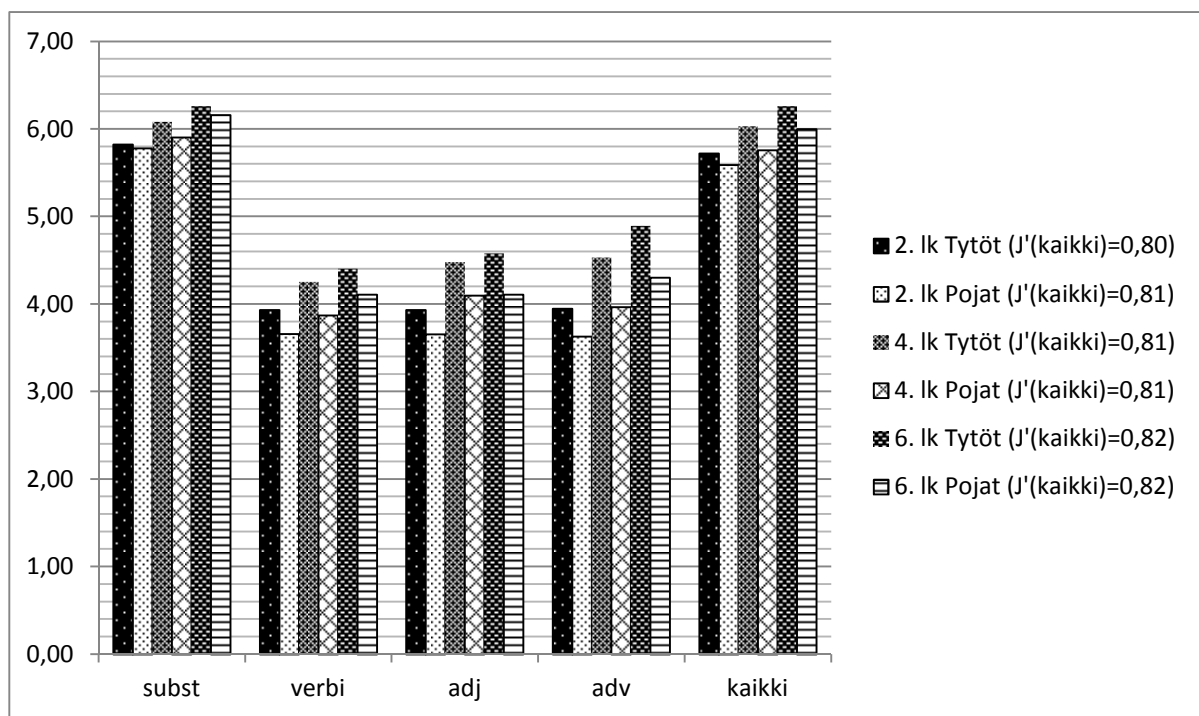
missä L on lekseemien määrä ja S saneiden määrä aineistossa.

Vielä tähän tutkimukseen otetaan mukaan Honorén R-indeksi, jota ainakin Mustonen (1997: 42) mukaan pidetään varsin luotettavana sana-sane-suhteen mittarina suhteellisen pienissäkin aineistoissa. R-indeksi lasketaan kaavalla

$$R = 100 \log \left(\frac{N}{1 - \frac{V_1}{V}} \right).$$

Kaavassa N on saneiden, V_1 HL-sanojen ja V lekseemien määrä.

Tässä tutkimuksessa sanaston monipuolisuutta tutkitaan rikkauslukujen avulla 2.-, 4.- ja 6.-luokkalaisten kirjoitelmista, sillä joka toisen luokan käyttäminen antaa todennäköisesti riittävän hyvän kuvan oppilaiden kirjoitustaidon kehittymisestä. Erikseen rikkauslukuja, tarkemmin Shannonin entropiaa, määritetään substantiiveista, verbeistä, adjektiiveista, adverbeista sekä koko sanastosta (kuvio 27). Muita, lähinnä vertailun vuoksi määritettäviä indeksejä, käytetään vain koko sanaston monipuolisuuden määrittämiseen. 2.-luokkalaisten laskelmissa käytetään koko aineistoa, 4.- ja 6.-luokkalaisten aineistossa vain sellaista aineistoa, jossa kirjoitelmille on annettu tasoluokitus. Numerot on poistettu tästä aineistosta, sillä niiden runsautta ei voi pitää sanastoa rikastuttavana tekijänä. Laskelmiin on käytetty rajattua aineistoa (ks. taulukko 18 s. 71).



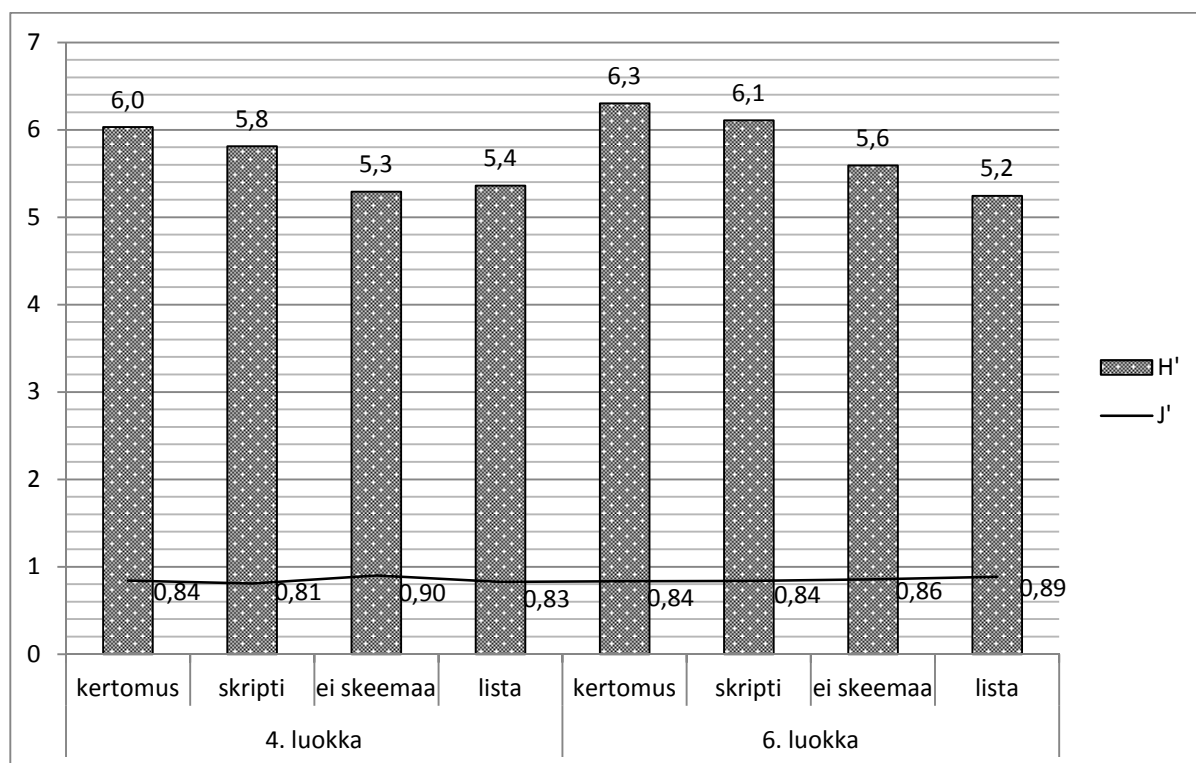
Kuvio 27. Shannonin entropia luokittain.

Shannonin entropian kaavalla määritetyt rikkausluvut (kuvio 27) kertovat selvästi tyttöjen sanaston olevan joka luokalla monipuolisempaa kuin poikien sanaston. Sanaston monipuolisuus näyttää kasvavan vuosi vuodelta – samalla kasvaa myös tyttöjen ja poikien välinen ero. Samansuuntaisia tuloksia on saanut myös Saarela (1991 ja 1997), joskin näyttää siltä, että oppilaiden sanasto on huomattavasti monipuolisempaa tässä aineistossa kuin Saarelan aineistossa. Tulos johtuu todennäköisesti siitä, että Saarelan tutkimuksessa mitataan

todellista kehitystä, eli kirjoittajina ovat aina samat oppilaat, kun taas tässä tutkimuksessa käytetään näennäisaikamenetelmää.

Jakauman tasaisuus on lähes sama kaikilla vertailuryhmillä. Saarelan (1997) saamiin tuloksiin verrattuna tämän tutkimuksen aineistossa sanastoa käytetään kuitenkin huomattavan epätasaisesti, sillä hänen aineistossaan jakauman J' keskiarvo on 2., 4. ja 6 - luokkalaisilla 0,95.

Seuraavaksi tarkastellaan sitä, minkälaista korrelaatiota kirjoitusten tason ja sanaston monipuolisuuden välille voidaan entropian avulla osoittaa (kuvio 28). 2.-luokkalaisten on jätetty laskelmien ulkopuolelle, sillä heidän aineistaan vain hyvin pienelle osalle on tehty tasoluokitus.



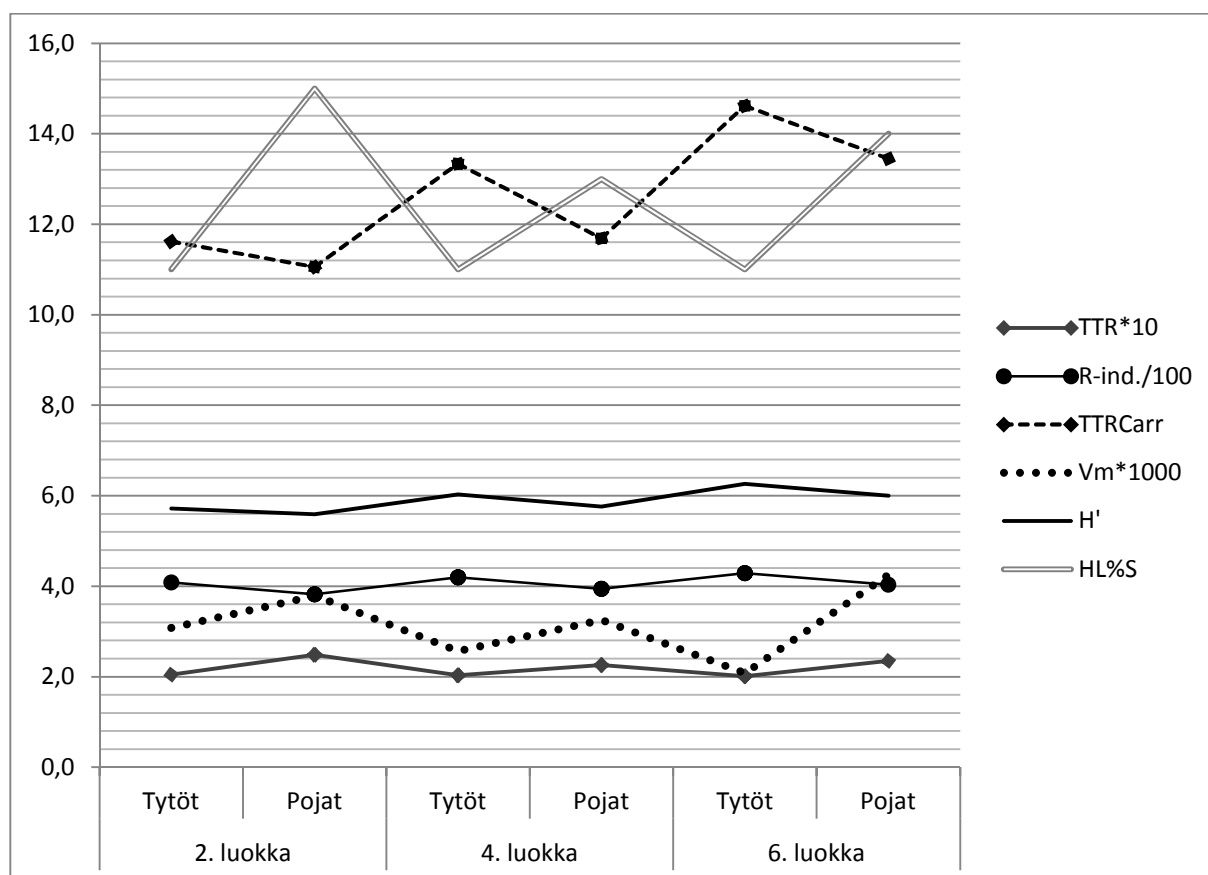
Kuvio 28. Koko sanaston monipuolisuus (H') ja jakauman tasaisuus (J') eritasoisissa kirjoitelmissa.

Eritasoisten kirjoitelmien sanaston vertailussa entropian suuruuteen vaikuttaa sekä kirjoitelman taso että kirjoitelmien määrä, jolla on yhteys tason koko sanemäärään. Pääpiirteittäin näyttää siltä, että kertomustasolla kirjoitelmien sanasto on huomattavasti monipuolisempaa kuin heikkotasoisempien kirjoitelmien sanasto. 4.-luokkalaisten kirjoitelmissa tämä toteutuu, vaikka skripti-tason kirjoitelmien yhteissanemäärä ja myös

erilaisten leksiemien määrä on suurempi kuin kertomustasolla. Muuten näyttää siltä, että entropiaan vaikuttaa huomattavasti kirjoitelmien sanemäärä ja tasaisuusluku taas on sitä suurempi, mitä vähemmän saneita koko tarkasteltavassa aineistossa on, vaikka aivan yksi yhteen tämä ei toteudu.

Jakauman tasaisuus ja taso näyttäisivät laskelmien perusteella korreloivan kuitenkin siten, että mitä heikkomuotoisempi kirjoitelma ja mitä suurempi entropia, sitä pienempi on jakauman tasaisuus. Toisaalta pieni sanemäärä ja pieni entropia näyttävät yleensä ennustavan suurempaa tasaisuutta. Tämä saattaa johtua siitä, että kirjoitelmissa käytetään paljon tiettyä aineiden aihepiiriin sopivaa perussanastoa.

Shannonin entropian lisäksi sanaston rikkauden mittareina on tässä tutkimuksessa käytetty TTR_{Carr} -indeksiä, Honorén R-indeksiä ja Herdanin V_m -indeksiä. Kaikkien käytettyjen mittareiden tulokset TTR-arvo ja HL%S mukaan lukien on koottu yhteen vertailun helpottamiseksi (kuvio 29). Mittareiden tuloksia on käsitelty siten, että ne sopivat hyvin samaan kuvaajaan. Mahdolliset jakajat tai kertoimet on esitetty seliteosassa.



Kuvio 29. Koonti rikkauslaskelmien tuloksista.

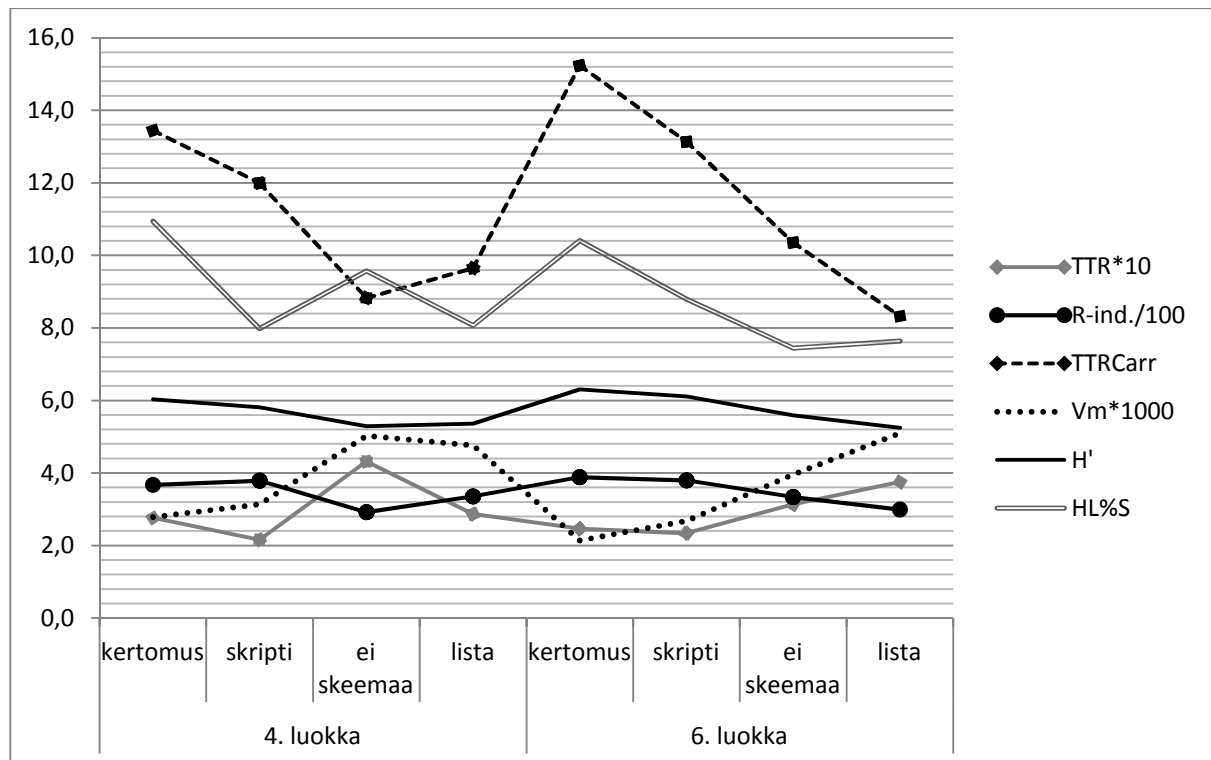
Kuvion perusteella tässä tutkimuksessa käytetyt rikkauslukulaskurit antavat samansuuntaisia tuloksia. Mittareista Shannonin entropia (H'), R-indeksi ja TTR_{Carr} antavat samanmuotoisen kuvaajan. V_m -indeksin kuvaajaa on luettava siten, että mitä pienempi on indeksin tulos, sitä rikkaampaa on kirjoitelmien sanasto.

Kaikki mittarit TTR-arvoa ja $HL\%S$ -arvoa lukuun ottamatta osoittavat, että tytöt käyttävät kaikilla vuosiluokilla sanastoa monipuolisemmin kuin pojat. Lähes kaikkien tässä käytettyjen indeksien mukaan 4. luokan tytöt näyttäisivät käyttävän sanastoa monipuolisemman kuin 6. luokan pojat. Erikoista V_m -indeksillä saaduissa tuloksissa on se, että 6.-luokkalaisten poikien sanasto näyttäisi olevan kaikkein yksipuolisinta - siis ”köyhempää” kuin esimerkiksi alempien luokkien poikien. Särkän (1987: 134) mukaan V_m - indeksi on yhtä lailla tyylien erottelija kuin sanaston runsauden mittari. Voisiko tässä olla syy 6.-luokkalaisten poikien ”köyhään” sanastonkäyttöön? Kenties V_m -mittarin tulokseen vaikuttavat 6.-luokkalaisten poikien väliset tasoerot, jotka ovat alkaneet kasvaa jälleen 4. luokan tasaantumisen jälkeen.

TTR-arvon perusteella poikien sanasto olisikin tyttöjä monipuolisempaa kaikilla vuosiluokilla ja $HL\%S$ -arvon mukaan 6.-luokkalaisten poikien sanaston käyttö olisi toiseksi rikkainta heti 2.-luokkalaisten poikien jälkeen, mikä tekee sekä TTR-arvon että $HL\%S$ -arvon antamista tuloksista erittäin kyseenalaisia. Kuten aiemmin on todettu, kirjoitelmien pituuden kasvaminen lisää toistoa, mikä pienentää $HL\%S$ -arvoa, ja tyttöjen kirjoitelmat ovat keskimäärin huomattavasti pidempiä kuin poikien.

$HL\%S$ -arvo ei näytä tytöillä muuttuvan lainkaan vuosien mittaan, kun taas poikien kohdalla arvo vaihtelee huomattavasti niin, että se on suurimmillaan 2. luokalla ja pienimmillään 4. luokalla. Kuitenkin kaikkien muiden mittarien antamien tulosten perusteella tyttöjen sanasto on aina poikien sanastoa monipuolisempaa ja 6.-luokkalaisten poikien sanasto on monipuolisempaa kuin alempien luokkien poikien sanasto, mikä vaikuttaa varsin järkeenkäyvältä tulokselta.

Lopuksi tarkastellaan rikkausindeksien tuloksia 4. ja 6. luokkien eritasoisissa kirjoitelmissa. Jo aiemmissa tutkimuksissa on saatu viitteitä siitä, että sanavaraston ja kirjoitelman tason välillä on yhteys (Matilainen 1989: 154), ja samaan viittaavat myös tässä tutkimuksessa saadut tulokset (kuvio 30).



Kuvio 30. Koonti rikkauslaskelmien tuloksista kirjoitusten tasojen mukaan.

Särkkä (1987: 135) toteaa, että rikas sanasto ei välttämättä kerro mitään kaunokirjallisen teoksen laadukkuudesta, ja Koulukorpus-aineistonkin tasoluokitukset on tehty ennen kaikkea skeeman koherenssin näkökulmasta. Tämän tutkimuksen aineistossa Särkänkin käyttämät HL%S ja V_m -arvojen suuruudet kuitenkin näyttävät kulkevan lähes joka tapauksessa käsi kädessä kirjoitelmien tason kanssa erityisesti 6. luokan aineistossa. Sen sijaan 4. luokan aineistossa ei skeemallisiksi arvioitujen tekstien tulokset ovat jatkuvasti huonommat kuin listaksi arvioitujen kirjoitelmien tulokset. Muihin indekseihin verrattuna HL%S-arvo ja etenkin TTR-arvo antavat edelleen poikkeavimpia tuloksia. HL%S-arvot eivät kuitenkaan ole yhtä lailla päinvastaisia kuin vuosiluokkien välisessä vertailussa.

4.- ja 6.-luokkalaisten välille rikkausluvut eivät tee selvää eroa. Parhaiden kirjoitelmien (kertomukset ja skriptit) suhteen TTR_{Carr} - ja V_m -indeksi sekä Shannonin entropia ovat ainoat, joiden perusteella 6.-luokkalaisten käyttämä sanasto olisi selvästi monipuolisempaa kuin 4.-luokkalaisten. 6.-luokkalaisten heikoimpien kirjoitelmien sanasto näyttäisi olevan usein jopa ”köyhempää” kuin 4.-luokkalaisten sanasto, mikä ei toki ole aivan mahdotonta, vaikka hieman yllättävää onkin. R-indeksin antamat tulokset ovat samansuuntaisia, mutta eivät lainkaan yhtä selkeitä kuin edellä mainittujen.

Monet kvantitatiivisia rikkausmittareita käyttäneet tutkijat (ks. esim. Särkkä 1987: 136, Mustonen 1997: 47, Jaakola 2004: 106–108) muistuttavat, että vaikka mittarit kertovatkin objektiivisia lukuja, sanaston monipuolisuuden arviointia ei voi jättää yksin niiden varaan. Ensinnäkin lukujen tulkinta ei ole yksiselitteistä, toiseksi on aina huomioitava tekijät, jotka lukuihin vaikuttavat, esimerkiksi otoksen koko ja mittareihin vaikuttavat tekijät kuten hajonnan huomioiminen tai huomiotta jättäminen, ja kolmanneksi: jo kielitiede tutkimusalana vaatii myös kvalitatiivista ja selittävää lähestymistapaa.

Rikkauslukujen vertailemisesta saattaisi tehdä luotettavampaa se, että kaikki aineistot olisivat tarpeeksi suuria ja samankokoisia. Tässä tutkimuksessa aineistona on kuitenkin aina käytetty tiettyjä kokonaisuuksia, jotka ovat sekä luokittain että tyttöjen ja poikien välillä erikokoisia. Suurempaan luotettavuuteen pääsemiseksi aineistojen olisi hyvä olla vähintään 10 000 sanetta, mutta mikäli verrataan tyttöjä, poikia tai eritasoisia aineita, 10 000 saneen otosta ei tämän tutkimuksen aineistosta ole mahdollista saada. Esimerkiksi 5000 tai 10 000 saneen poimiminen luokittaisistakin aineistoista olisi hieman kyseenalaista, sillä kirjoitelmien tasoerot ovat niin huomattavia, että parhaimpien tai heikoimpien aineiden painottuminen tuossa valinnassa saattaisi vaikuttaa tuloksiin merkittävästi.

4 YHTEENVETO

Tämän tutkimuksen tavoitteena on esitellä Koulukorpuksen leksikkoa yleisesti sekä vertailla leksikon kehittymistä alakoulun aikana niin eri vuosiluokkien kuin sukupuoltenkin välillä. Lienee selvää, että mitä vanhempien lasten kirjoitelmista on kyse, sitä monipuolisempaa sanastoa kirjoitelmista löytyy – ainakin luokkatasolla. Sekä lekseemien määrä per oppilas että rikkauslukulaskelmien tulokset vahvistavatkin tämän hypoteesin. Toinen hypoteesi on, että tytöt käyttävät monipuolisempaa sanastoa kuin pojat. Tämänkin hypoteesin erityisesti rikkauslukulaskelmat varsin selkeästi vahvistavat.

Rikkauslukujen valossa kirjoitelmien taso näyttäisi korreloivan varsin hyvin leksikon monipuolisuuden kanssa. Samalla kun tulokset osoittavat, että hyvätasoisten kirjoitelmien sanasto on yleensä huonotasoisempien kirjoitelmien sanastoa monipuolisempaa, ne myös vahvistavat alustavien tasoluokitusten onnistuneisuutta. Loppujen lopuksihan kirjoitelmien taso on määritetty pitkälti tutkijoiden intuition pohjalta. Vaikka rikkauslukulaskelmiin sanaston monipuolisuuden määrittelyssä pitääkin aina suhtautua varauksella, näyttävät ne tässä tutkimuksessa toimivan ainakin hypoteesien vahvistajina.

Aineisto on koottu 1.–6.-luokkalaisten Unelmien(i) päivä -aiheisista kirjoitelmista. Koko aineiston kymmenen frekventeintä lekseemiä ovat *olla, ja, se, minä, mennä, kun, sitten, päivä, me* ja *saada*. Nämä kymmenen lekseemiä kattavat koko aineiston saneista yhteensä runsaan neljänneksen. Aineiston kymmenen yleisintä substantiivia ovat *päivä, unelma, äiti, koti, kaveri, aamu, koira, isä, koulu* ja *ilta*, jotka kattavat yhteensä 16 % kaikista aineiston substantiivisaneista. Yleisimmät kymmenen verbiä taas kattavat jopa hieman yli puolet kaikista verbisaneista. Ne ovat *olla, mennä, saada, ei, lähteä, tulla, syödä, haluta, ostaa* ja *käydä*. Tytöt ja pojat käyttävät pitkälti samoja peruslekseemejä, mutta harvemmin käytetyistä lekseemeistä suuri osa on sellaisia, joita tytöt käyttävät, mutta pojat eivät.

Koko aineiston yleisimmät lekseemit ovat varsin konkreettisia ja lyhyitä. Ainoa tavallista abstraktimpi lekseemi, *unelma*, selittyy kirjoitelmien aiheella. Vaikka otsikot onkin poistettu aineistosta, suoraan aiheeseen liittyvät lekseemit ovat silti hyvin frekventtejä, sillä moni kirjoittaja on toistanut aiheen kirjoitelmassaan.

Koulukorpuksen leksikkoa luonnehtii erityisen hyvin yleisimpien lekseemien kumulatiivinen frekvenssi. Verrattaessa Koulukorpus-aineiston kumulatiivista frekvenssiä oppikirjatekstien ja sanomalehtikielen sanastoihin näkyy, että Koulukorpus-aineistossa yleisimmät lekseemit

kattavat huomattavan ison osan saneista. Kun Koulukorpuksessa sata yleisintä lekseemiä kattaa saneista yli 55 %, yli 50 %:n kattavuuteen päästään sanomalehtikielen taajuussanastossa vasta tuhannen yleisimmän lekseemin kohdalla. Toki tätä tulosta selittää paljon koululaisten kirjoitelmien yhteinen otsikko ja Koulukorpus-aineiston vertailuaineistoja pienempi koko, mutta osin selitys löytyy myös alakouluikäisten suppeammasta sanavarastosta. Mitä ylemmille luokille siirrytään, sitä pienemmän osan yleisimmät lekseemit kuitenkin vuosiluokan aineistosta kattavat.

Saneiden sanaluokkien osuuksissa ei prosenttiyksikköjen valossa näytä tapahtuvan kovin suuria muutoksia 1. luokalta 6. luokalle, mutta yhdenkin prosenttiyksikön muutokset voivat kertoa jotakin olennaista kirjoitustaidon kehittymisestä. Substantiivien osuus vaihtelee, joskin niiden osuus näyttäisi hieman kasvavan ylemmillä luokilla. Verbisaneiden osuus hieman pienenee, kun taas esimerkiksi kuvailuun käytettävien adjektiivien ja adverbien osuudet kasvavat. Pronominien osuus laskee hieman, vaikka erilaisia pronominileksemejä käytetäänkin monipuolisemmin. Interjektiot, konjunktiot ja adpositiot olen niputtanut ryhmään ”muut partikkelit”. Niiden osuus pienenee lasten varttuessa. Syynä tähän voi olla se, että nuoret oppilaat saattavat usein käyttää kirjoitelmissaan enemmän dialogia, jossa interjektioille on enemmän käyttöä. Vanhemmat lapset taas osaavat hyödyntää enemmän referointia ja epäsuoraa esitystä, johon dialogipartikkelit tai interjektiot eivät yhtä hyvin sovi.

Lekseemien sanaluokkien osuuksissa tapahtuvat muutokset luokka-asteilta ylemmille siirryttäessä eroavat hieman saneiden sanaluokissa tapahtuvista muutoksista. Substantiivien osuus pienenee, ennen kuin se jälleen 6. luokalla kasvaa. Verbien osuus taas kasvaa alakoulun aikana jatkuvasti samoin kuin adverbien ja adjektiivien osuudet, mikä kertoo erityisesti verbien valikoiman kasvamisesta ja ehkä myös kirjoitustaidon kehittymisestä, johon verbien, adjektiivien ja adverbien monipuolisempi käyttö usein liitetään. Suuri substantiivien osuus ainakin alimmilla vuosiluokilla johtuu usein pitkistä substantiiviluetteloista, jotka eivät juuri vie kertomusta eteenpäin. Pronominilekseemien osuus pienenee ylemmillä vuosiluokilla, mikä on luonnollista. Suljettuna sanaluokkana pronominiinien osuus ei voi kasvaa loputtomiin, vaikka pronominileksemejä otettaisiinkin käyttöön enemmän.

Olen analyysissäni verrannut Koulukorpuksen leksikkoon muutamiin muihin aineistoihin tulosten suhteuttamiseksi, sillä pelkkien sanalistojen esittäminen ei ole erityisen mielekäs eikä auta välttämättä ymmärtämään alakoululaisten sanaston erityispiirteitä. Vertailuaineistot

ovat olleet CSC:n internet-sivuilla löytyvä Suomen sanomalehtikielen taajuussanasto (2004), Opettajan sanastokirja (1970), Jaakolan (2004) pro gradu -tutkielma 2. ja 3. -luokkalaisten oppikirjojen sanaston dynamiikasta sekä Peltosen (2011) pro gradu -tutkielma Substantiivit, adjektiivit ja idiomaattiset ilmaukset lasten kuvakirjoissa. Vertailua tehdessäni olen toki tiedostanut, että koska sekä kouluaine tekstilajina että aineen aihepiiri vaikuttavat sanastoon ja kieleen merkittävästi, hedelmällisintä vertailu on Koulukorpus-aineiston sisällä eri vuosiluokkien ja sukupuolten välillä. Vertailu muihin aineistoihin kuitenkin osoittaa, että koululaiset käyttävät paljon perussanastoa ja että kirjoittajien taidot, tekstilaji ja aineen aihe ovat vaikuttaneet sanastoon merkittävästi. Esimerkiksi sana *unelma*, joka on Koulukorpus-aineistossa 20. yleisin lekseemi, löytyy CSC:n Suomen Sanomalehtikielen taajuussanastosta vasta 2334. sijalta ja Opettajan sanastokirjan tulosten mukaan kyseistä lekseemiä on käyttänyt vain yksi kirjoittaja tuhannesta. Aihe ja tekstilaji vaikuttavat *minä*-lekseemin frekvenssiin, sillä unelmien päivästä kirjoitetaan usein minä-muodossa. *minä*-lekseemin sijoittuminen neljännelle sijalle kertoo kirjoitustaidoista sen, että nuoret kirjoittajat käyttävät kyseistä pronominia paljon silloinkin, kun sen voisi jättää pois. *minä*-lekseemin käyttö väheneekin ylemmille vuosiluokille siirryttäessä.

Alakoulun aikaista kehitystä tutkiakseni olen useasti vertaillut keskenään vuosiluokkia 2, 4 ja 6. Niiden välisessä vertailussa 4. luokka poikkeaa usein jollakin lailla 2. ja 6. luokan trendistä. Yhtenä syynä voi olla 4. luokan aineiston heterogeenisuus, sillä se sisältää kielikylpykoululaisten kirjoitelmia, mutta toisaalta on tuskin luonnotonta, jos sanaston kehitys ei yleensääkään tapahdu tasaisesti ja lineaarisesti. 4. luokan aineiston heterogeenisuudesta huolimatta tai juuri siksi näyttää siltä, että 4. luokalla poikien ja tyttöjen väliset erot ovat pienimmillään ja että 6. luokalla tasoerot alkavat jälleen kasvaa. Syynä kasvaviin eroihin voi olla sekin, että joillakin oppilailla murrosikä vaikuttaa jo häiritsevästi koulunkäyntiin.

Kaikkien lasten sanavarasto karttuu koulussa paljon, mutta jotkut omaksuvat tietoa ja sanoja huomattavasti toisia paremmin. Myös vapaa-ajan harrastukset eriytyvät todennäköisesti yhä enemmän lasten varttuessa, mikä näkyy ainakin eri harrastuksiin liittyvien erityissanojen hallintana. Kun erilaiset harrastukset eriyttävät sanavarastoa, on lukuharrastuksella sanavarastoon yleisesti rikastuttava vaikutus. On havaittu, että ahkera vapaa-ajalla lukeminen vahvistaa äidinkielen ja kirjallisuuden eri osa-alueiden hallintaa (Lappalainen 2008: 6). Lukeminen on enemmän tyttöjen kuin poikien harrastus, minkä lisäksi kulttuurilliset tekijät saattavat vaikuttaa siihen, että tytöt kirjoittavat paremmin ja käyttävät rikkaampaa kieltä kuin pojat. Kirjoittamista kun pidetään ehkä enemmän tyttömäisenä kuin

poikamaisena harrastuksena. Rikkauslukujen antamien tulosten valossa näyttää siltä, että poikien sanavaraston kasvu taittuu kuudennella luokalla, mikä kasvattaa ennestään railoa tyttöjen ja poikien välillä. Yksilöllinen variaatio on toki suurta sekä tyttöjen että poikien välillä.

Olen pyrkinyt tutkimaan ja kuvailemaan Koulukorpuksen leksikkoa kvantitatiivisesti ja hieman myös kvalitatiivisesti siten, että alakouluikäisten sanaston olennaisia piirteitä tulisi näkyviin. Eri sanaluokkia on esimerkiksi tutkittu jonkin verran erilaisista näkökulmista riippuen siitä, mitkä piirteet ovat kussakin tapauksessa nousseet aineistosta esiin. Syvempää analyysia sanaston ja kouluikäisten kielen eri osioista vaikkapa yksittäisten sanaluokkien tai syntaksin osalta on tarjolla useissa muissa, esimerkiksi johdantokappaleessa mainitsemissani tutkielmissa, joissa Koulukorpusta on käytetty aineistona.

Tämän tutkimuksen tulosten tarkastelussa on aina huomioitava se, että koko aineistossa alimpien luokkien saneiden osuus on huomattavasti pienempi kuin ylimpien. Kun 1.-luokkalaisten aineisto käsittää vajaat 3 000 sanetta, lähestyvät 4.–6.-luokkien aineistot 20 000 sanetta. 10–12-vuotiaiden sanaston osuus on siis koko aineiston tuloksissa vallitseva (n. 68 % aineistosta). Sukupuolten välinen jakauma on saneiden osalta hyvin epätasainen. Vaikka tyttökirjoittajia onkin ollut vain 15 enemmän kuin poikakirjoittajia, tyttöjen kirjoitelmien yhteissanemäärä on lähes 50 000 poikien sanemäärän jäädessä runsaaseen 32 000:een. Tyttöjen kirjoitelmien yhteissanemäärä on siis yli 1,5-kertainen poikien kirjoitelmien sanemäärään verrattuna. Aineiston koko vaikuttaa aina jollakin tavalla kvantitatiivisiin tuloksiin, sillä kuten on jo todettu, esimerkiksi HL-sanojen osuus saneista on yleensä sitä pienempi mitä suurempi aineisto on. Suurista aineistoista saadut rikkauslukulaskelmien tulokset ovat myös todennäköisesti tarkempia kuin pienistä aineistoista saadut tulokset. Marginaalisesti tuloksiin saattaa vaikuttaa myös aineiston koodauksen horjuvuus, mutta merkittävää se ei tämän tutkimuksen tulosten kannalta ole.

Kirjoitelmien aiheen vaikutusta sanastoon ei voi väheksyä. Se näkyy selvästi esimerkiksi yleisimpien substantiivien listalla, joka sisältää muun muassa lekseemit *unelma*, *Särkänniemi* ja *Lotto* sekä yleisimpien verbien listalla, johon kuuluvat muun muassa matkustamiseen liittyvät verbit *lentää*, *pakata* ja *matkustaa*. Toisaalta se, että kaikki tutkimusaineistoa tuottaneet 1.–6. luokan oppilaat ovat kirjoittaneet samasta aiheesta, mahdollistaa sanastossa alakoulun aikana tapahtuvien muutosten havainnoinnin niin frekvenssien, sanaston monipuolisuuden kuin yleisimpien lekseemienkin näkökulmasta. Kun

tutkimuksen abstraktiotasoa nostetaan pintatason lekseemien tutkimuksesta vaikkapa niiden semanttiseen tarkasteluun, aiheen vaikutus tuloksiin todennäköisesti vähenee.

Koulukorpuksen leksikon yleinen tarkastelu olisi tarjonnut useampiakin mielenkiintoisia tutkimuskohteita kuin mihin tämän tutkielman puitteissa on voitu tarttua. Esimerkiksi HL-sanojen semanttinen tai muuten yksityiskohtaisempi tarkastelu ja luokkien välinen vertailu olisi antanut mielenkiintoista tietoa alakoululaisten HL-sanojen luonteesta. Myös kirjoitelmien tasoluokituksia olisi voinut hyödyntää vertailussa runsaammin, mutta se on vielä hedelmällisempää sen jälkeen, kun tasoluokituskriteerit on tarkistettu ja kaikki aineiston kirjoitelmat on luokiteltu.

Tutkimuksen lopuksi voi todeta, että alakouluikäisten aktiivisen sanavaraston lisäksi kirjoitelmien yleisimmät lekseemit kertovat paljon myös ajastamme ja lasten vapaa-ajanviettotavoista nykyään. Puhekielinen *shoppailla* tai *tietokone* eivät todennäköisesti olisi olleet yhtä frekventtejä lasten kirjoitelmissa 40 vuotta sitten, vaikka kirjoitelmien aihe olisikin ollut sama.

Lähteet

Aineistolähde

PAJUNEN, ANNELI – PURANEN, MAIJA – YLI-PAAVOLA, ANNUKKA 2010: *Koulukorpus 2007–2011*. Opiskelijaversio. Suomen kieli. Tampereen yliopisto.

Kirjallisuuslähteet

BERMAN, RUTH A. 2004: Between emergence and mastery. The long developmental route of language acquisition. *Language Development Across Childhood and Adolescence* (Berman, Ruth Aronson, toim.) Amsterdam, NLD: John Benjamins Publishing Company.

DOCKRELL, JULIE E. – MESSER, DAVID 2004: Lexical acquisition in the early school years. *Language Development Across Childhood and Adolescence*. Berman, Ruth Aronson (toim.) Amsterdam, NLD: John Benjamins Publishing Company.

JAAKOLA, MARI 2004: *Peruskoulun 2. ja 3. luokan oppikirjojen sanasto – dynamiikasta luettavuuteen*. Pro gradu -tutkielma. Tampereen yliopisto, suomen kieli.

KARLSSON, FRED 2003 (4. uud. painos (1994)): *Yleinen kielitiede*. Helsinki: Yliopistopaino.

KOIVUNEN, JUSSI 2012: *Adverbien käyttö alakoululaisten kirjoitelmissa*. Pro gradu -tutkielma. Tampereen yliopiston kieli-, kirjallisuus- ja käännöstieteiden yksikkö.

KOPPINEN, MAIJA-LEENA 1982: Kielellisen ilmaisun aktivoitumisesta. – Maija Larmola (toim.), *Kouluikäisten kieli* s. 72–109. Tietolipas 88. Helsinki: Suomalaisen Kirjallisuuden Seura.

KUISMA, JEMINA 2011: *Ominaisuuden ilmausten käytön kehittyminen alakouluikäisten kirjoitelmissa*. Tampereen yliopiston kieli- ja käännöstieteiden laitos.

LAPPALAINEN, HANNU-PEKKA 2008: *On annettu hyviä numeroita*. Perusopetuksen 6. vuosiluokan suorittaneiden äidinkielen ja kirjallisuuden oppimistulosten arviointi 2007. Oppimistulosten arviointi 3/2008. Helsinki: Opetushallitus.

LEINO, PIRKKO 1989: *Suomen kielioppi*. Helsinki: Otava.

LEIWO, MATTI 1980: *Lapsen kielen kehitys*. Helsinki: Gaudeamus.

MATILAINEN, KAIJA 1989: *Kirjoitustaidon kehittyminen neljän ensimmäisen kouluvuoden aikana*. Joensuun yliopiston kasvatustieteellisiä julkaisuja. No 9. Joensuu: Joensuun yliopiston monistuskeskus.

MUSTONEN, HARRI 1997: *Mitä sanasto kertoo maailmankuvasta? Tutkimusmenetelmien kehittelyä sekä yhden puhujan analyysi*. Pro gradu -tutkielma. Tampereen yliopiston kasvatustieteiden ja kielitieteen osasto.

NIEMIKORPI, ANTERO 1991: *Suomen kielen sanaston dynamiikkaa*. Acta Wasaensia. No 26. Kielitiede 2. Vaasa: Vaasan yliopistopaino.

NIPPOLD, MARILYN A. 2004: Research on later language development. International perspectives. *Language Development Across Childhood and Adolescence* (Berman, Ruth Aronson, toim.) Amsterdam, NLD: John Benjamins Publishing Company.

OPETTAJAN SANASTOKIRJA = KARVONEN, JUHANI – RÖMÄN, KYLLIKKI – TAKALA, ANNIKA – YLINENTALO OIVA 1970: *Opettajan sanastokirja*. Jyväskylä: Gummerus.

PAJUNEN, ANNELI 2012: Kirjoittamistaitojen kehitys 8–12-vuotiailla. Alakoululaisten unelmakirjoitelmat. –*Virittäjä* 116 s. 4–32.

PELTONEN, JUTTA 2011: *Substantiivit, adjektiivit ja idiomaattiset ilmaukset lasten kuvakirjoissa*. Pro gradu -tutkielma. Tampereen yliopiston kieli- ja kirjallisuustieteiden yksikkö.

PIAGET, JEAN 1973: *Språk och tanke hos barnet*. Lund: Bröderna Ekstrands Tryckeri AB.

PIAGET, JEAN 1988: *Lapsi maailmansa rakentajana. Kuusi esseitä lapsen kehityksestä*. Juva: WSOY.

- SAARELA, LEENA 1991: *Neljäsluokkalaisten tyttöjen ja poikien sanaston eroista*. Oulun yliopiston suomen ja saamen kielen laitoksen tutkimusraportteja 33. Oulu: Oulun yliopisto.
- SAARELA, LEENA 1997: *Peruskoululaisten kirjoitelmien kehittyminen sanastontutkimuksen valossa*. Acta Universitatis Ouluensis. B Humaniora 25. Oulu: Oulun yliopisto.
- SAARINEN, PIRKKO 1982: Kouluikäisten kielellinen luovuus. – Maija Larmola (toim.), *Kouluikäisten kieli* s. 55–71. Tietolipas 88. Helsinki: Suomalaisen Kirjallisuuden Seura.
- SINTONEN, MARLEENA 2011: *Suunnittelemisen vaikutus kertomuksen rakenteeseen viides- ja seitsemäsluokkalaisilla kirjoittajilla*. Pro gradu -tutkielma. Tampereen yliopiston kieli- ja käännöstieteiden laitos.
- SÄRKKÄ, TAUNO 1987: Sanaston rikkaudesta ja sen mittaamisesta. –*Virittäjä* 91 s. 129–136.
- TOLCHINSKY, LILIANA 2004: The nature and scope of later language development. *Language Development Across Childhood and Adolescence* (Berman, Ruth Aronson, toim.) Amsterdam, NLD: John Benjamins Publishing Company.

Verkkolähteet

- POPS2004 = OPETUSHALLITUS 2004: Perusopetuksen opetussuunnitelman perusteet. - http://www.oph.fi/koulutuksen_jarjestaminen/opetussuunnitelmien_ja_tutkintojen_perusteet/perusopetus. 25.6.2012
- Tieteen tietotekniikan keskus CSC: Suomen sanomalehtikielen taajuussanasto (koottu 2004). Viitattu 1.4.2012. Saatavissa: <http://www.csc.fi/tutkimus/alat/kielitiede/taajuussanasto-B9996/view>. 1.4. 2012
- VISK = AULI HAKULINEN, MARIA VILKUNA, RIITTA KORHONEN, VESA KOIVISTO, TARJA RIITTA HEINONEN JA IRJA ALHO 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio, viitattu 1.11.2008. Saatavissa: <http://scripta.kotus.fi/visk> URN:ISBN:978-952-5446-35-7. 1.4.2012

Liite 1: Numeerisia tietoja aineistosta

| | 1. luokka | 2. luokka | 3. luokka | 4. luokka | 5. luokka | 6. luokka | koko aineisto |
|---|-----------|-----------|-----------|-----------|-----------|-----------|---------------|
| oppilaita | 123 | 224 | 207 | 215 | 182 | 160 | 1111 |
| tyttöjä | 63 | 115 | 99 | 118 | 90 | 78 | 563 |
| poikia | 60 | 109 | 108 | 97 | 92 | 82 | 548 |
| saneita yhteensä noin | 2700 | 10500 | 13400 | 18800 | 19400 | 17600 | 82300 |
| saneita tyttöjen aineissa yhteensä | 1785 | 6475 | 7396 | 11942 | 11586 | 10729 | 49913 |
| saneita poikien aineissa yhteensä | 912 | 3992 | 5932 | 6889 | 7803 | 6857 | 32385 |
| lekseimejä yhteensä noin | 730 | 1880 | 2440 | 3020 | 2955 | 3000 | 7770 |
| aineiden pituus keskimäärin (sanetta) | 22 | 47 | 66 | 88 | 107 | 110 | |
| aineiden pituuden keskihajonta | 18 | 26 | 39 | 46 | 56 | 59 | |
| pisin aine (sanetta) | 94 | 134 | 236 | 233 | 329 | 346 | |
| lyhin aine (sanetta) | 1 | 4 | 5 | 8 | 17 | 6 | |
| SANEET | | | | | | | |
| substantiiveja | 864 | 3189 | 3799 | 5607 | 5953 | 5356 | 24788 |
| verbejä | 790 | 3049 | 3741 | 5521 | 5682 | 5000 | 23783 |
| adjektiiveja | 120 | 464 | 743 | 949 | 984 | 1022 | 4282 |
| adverbeja | 220 | 1034 | 1518 | 2025 | 2106 | 2064 | 8967 |
| pronomineja | 224 | 1202 | 1517 | 1901 | 1956 | 1672 | 8472 |
| numeraaleja (sisältää numerot) | 24 | 130 | 271 | 450 | 384 | 371 | 1643 |
| <i>numeraalit (ilman numeroita)</i> | 13 | 79 | 173 | 260 | 237 | 209 | 971 |
| interjektioita | 3 | 56 | 28 | 54 | 94 | 39 | 274 |
| pre/postpositioita | 41 | 178 | 215 | 352 | 283 | 243 | 1312 |
| konjunktioita | 384 | 1162 | 1482 | 1964 | 1920 | 1782 | 8694 |
| muut (ongelmalliset) | 27 | 2 | 5 | 6 | 27 | 16 | 67 |
| substantiivisaneiden ja verbisaneiden suhde | 1,10 | 1,05 | 1,03 | 1,02 | 1,05 | 1,07 | 1,04 |
| LEKSEEMIT | | | | | | | |
| substantiiveja | 429 | 1104 | 1376 | 1672 | 1736 | 1746 | 5001 |
| verbejä | 113 | 303 | 399 | 548 | 574 | 550 | 1167 |
| adjektiiveja | 53 | 130 | 197 | 233 | 262 | 247 | 614 |
| adverbeja | 63 | 170 | 238 | 276 | 357 | 316 | 629 |
| pronomineja | 20 | 35 | 40 | 37 | 34 | 36 | 60 |
| <i>numeraalit (ilman numeroita)</i> | 9 | 23 | 38 | 35 | 32 | 29 | 67 |
| interjektioita | 3 | 31 | 13 | 23 | 44 | 20 | 93 |
| pre/postpositioita | 13 | 37 | 43 | 35 | 22 | 7 | 90 |
| konjunktioita | 12 | 19 | 22 | 21 | 24 | 21 | 36 |
| ? | 27 | 2 | 5 | 3 | 9 | 0 | 37 |
| substantiivilekseemien suhde saneiden määrään | 0,16 | 0,11 | 0,10 | 0,09 | 0,09 | 0,10 | 0,06 |
| verbilekseemien suhde saneiden määrään | 0,04 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,01 |
| yhden kerran esiintyviä lekseimejä sis. numerot | 472 | 1031 | 1371 | 1629 | 1768 | 1607 | 4241 |
| yhden kerran esiintyviä lekseimejä ei sis. nroita | 463 | 1001 | 1312 | 1513 | 1693 | 1611 | 4023 |
| yhden kerran esiintyviä lekseimejä % saneista | 18 | 10 | 10 | 9 | 9 | 9 | 5 |
| yhden kerran esiintyviä lekseimejä % lekseemeistä | 62 | 55 | 56 | 54 | 55 | 55 | 54 |

Liite 2: Koko aineiston yleisimmät lekseemit (n = 100)

| <u>lekseemi</u> | <u>kpl</u> | <u>% saneista</u> | | | |
|-----------------|------------|-------------------|----------|-----|------|
| olla | 5352 | 6,50 | aika | 225 | 0,27 |
| ja | 4827 | 5,86 | pitää | 214 | 0,26 |
| se | 1893 | 2,30 | jo | 210 | 0,26 |
| minä | 1867 | 2,27 | ne | 209 | 0,25 |
| mennä | 1826 | 2,22 | hyvä | 208 | 0,25 |
| kun | 1308 | 1,59 | ilta | 208 | 0,25 |
| sitten | 1093 | 1,33 | oma | 198 | 0,24 |
| päivä | 1070 | 1,30 | koko | 190 | 0,23 |
| me | 1042 | 1,27 | ottaa | 189 | 0,23 |
| saada | 1006 | 1,22 | kello | 187 | 0,23 |
| että | 948 | 1,15 | tunti | 185 | 0,22 |
| ei | 806 | 0,98 | he | 183 | 0,22 |
| lähteä | 767 | 0,93 | yksi | 180 | 0,22 |
| tulla | 750 | 0,91 | raha | 178 | 0,22 |
| joka | 603 | 0,73 | tosi | 176 | 0,21 |
| syödä | 572 | 0,69 | iso | 174 | 0,21 |
| niin | 551 | 0,67 | jos | 173 | 0,21 |
| kanssa | 538 | 0,65 | paras | 169 | 0,21 |
| mutta | 509 | 0,62 | uusi | 168 | 0,20 |
| unelma | 497 | 0,60 | voittaa | 167 | 0,20 |
| äiti | 495 | 0,60 | matka | 162 | 0,20 |
| siellä | 476 | 0,58 | nimi | 162 | 0,20 |
| kaikki | 471 | 0,57 | nähdä | 162 | 0,20 |
| haluta | 467 | 0,57 | hevonen | 159 | 0,19 |
| hän | 448 | 0,54 | nyt | 151 | 0,18 |
| ostaa | 396 | 0,48 | auto | 149 | 0,18 |
| jälkeen | 395 | 0,48 | hotelli | 148 | 0,18 |
| päästä | 385 | 0,47 | heti | 147 | 0,18 |
| koti | 372 | 0,45 | tai | 147 | 0,18 |
| käydä | 355 | 0,43 | loppu | 145 | 0,18 |
| voida | 336 | 0,41 | kuin | 143 | 0,17 |
| kaveri | 322 | 0,39 | jokin | 141 | 0,17 |
| sanoa | 316 | 0,38 | ihana | 140 | 0,17 |
| nukkua | 308 | 0,37 | aamupala | 139 | 0,17 |
| paljon | 307 | 0,37 | muu | 139 | 0,17 |
| aamu | 299 | 0,36 | toinen | 136 | 0,17 |
| herätä | 298 | 0,36 | hauska | 134 | 0,16 |
| mikä | 297 | 0,36 | karkki | 133 | 0,16 |
| uida | 270 | 0,33 | kaksi | 132 | 0,16 |
| koira | 269 | 0,33 | tämä | 129 | 0,16 |
| koska | 269 | 0,33 | vain | 127 | 0,15 |
| katsoa | 263 | 0,32 | hieno | 126 | 0,15 |
| tehdä | 263 | 0,32 | ruoka | 126 | 0,15 |
| isä | 260 | 0,32 | vaikka | 126 | 0,15 |
| pelata | 246 | 0,30 | seuraava | 120 | 0,15 |
| alkaa | 243 | 0,30 | kysyä | 119 | 0,14 |
| kiva | 237 | 0,29 | leikkiä | 119 | 0,14 |
| koulu | 232 | 0,28 | toivoa | 119 | 0,14 |
| myös | 230 | 0,28 | hakea | 117 | 0,14 |
| vielä | 226 | 0,27 | ystävä | 116 | 0,14 |

Liite 3: Koko aineiston yleisimmät substantiivit (n=102)

| <u>lekseemi</u> | <u>kpl</u> | <u>% substantiiveista</u> | | | |
|-----------------|------------|---------------------------|-------------|----|-----|
| päivä | 1068 | 4,3 | tietokone | 67 | 0,3 |
| unelma | 497 | 2,0 | puoli | 66 | 0,3 |
| äiti | 495 | 2,0 | viikko | 66 | 0,3 |
| koti | 342 | 1,4 | elokuva | 64 | 0,3 |
| kaveri | 322 | 1,3 | mökki | 63 | 0,3 |
| aamu | 299 | 1,2 | kesä | 59 | 0,2 |
| koira | 269 | 1,1 | sisko | 59 | 0,2 |
| isä | 260 | 1,0 | uima-allas | 59 | 0,2 |
| koulu | 231 | 0,9 | kerta | 57 | 0,2 |
| ilta | 208 | 0,8 | koiranpentu | 56 | 0,2 |
| kello | 187 | 0,8 | lentokone | 56 | 0,2 |
| tunti | 185 | 0,7 | Särkänniemi | 56 | 0,2 |
| aika | 180 | 0,7 | ovi | 55 | 0,2 |
| raha | 178 | 0,7 | jäätelö | 54 | 0,2 |
| matka | 162 | 0,7 | laite | 54 | 0,2 |
| nimi | 162 | 0,7 | eläin | 53 | 0,2 |
| hevonen | 159 | 0,6 | ravintola | 53 | 0,2 |
| auto | 149 | 0,6 | aamiainen | 52 | 0,2 |
| hotelli | 148 | 0,6 | hetki | 52 | 0,2 |
| loppu | 143 | 0,6 | ihminen | 52 | 0,2 |
| aamupala | 139 | 0,6 | kala | 52 | 0,2 |
| karkki | 132 | 0,5 | esimerkki | 51 | 0,2 |
| ruoka | 126 | 0,5 | loma | 51 | 0,2 |
| ystävä | 116 | 0,5 | poika | 50 | 0,2 |
| kissa | 115 | 0,5 | poni | 48 | 0,2 |
| yö | 113 | 0,5 | vanhemmat | 47 | 0,2 |
| pele | 110 | 0,4 | pentu | 45 | 0,2 |
| perhe | 110 | 0,4 | sauna | 45 | 0,2 |
| huone | 107 | 0,4 | meri | 44 | 0,2 |
| maailma | 102 | 0,4 | mieli | 44 | 0,2 |
| lahja | 99 | 0,4 | palvelija | 44 | 0,2 |
| kauppa | 96 | 0,4 | laiva | 43 | 0,2 |
| euro | 92 | 0,4 | maali | 43 | 0,2 |
| sänky | 92 | 0,4 | metsä | 43 | 0,2 |
| talo | 89 | 0,4 | asia | 42 | 0,2 |
| tyttö | 89 | 0,4 | bussi | 42 | 0,2 |
| aurinko | 87 | 0,4 | synttär | 42 | 0,2 |
| vaate | 86 | 0,3 | tavara | 42 | 0,2 |
| vesi | 86 | 0,3 | pari | 41 | 0,2 |
| piha | 81 | 0,3 | elämä | 40 | 0,2 |
| Helsinki | 74 | 0,3 | iskä | 40 | 0,2 |
| Suomi | 74 | 0,3 | kesäloma | 40 | 0,2 |
| talli | 74 | 0,3 | klo | 40 | 0,2 |
| vuosi | 73 | 0,3 | Lotto | 40 | 0,2 |
| paikka | 72 | 0,3 | uni | 40 | 0,2 |
| ranta | 72 | 0,3 | iltapala | 39 | 0,2 |
| kirja | 71 | 0,3 | mies | 39 | 0,2 |
| syntymäpäivä | 70 | 0,3 | allas | 38 | 0,2 |
| kaupunki | 68 | 0,3 | huvipuisto | 38 | 0,2 |
| maa | 67 | 0,3 | keittiö | 38 | 0,2 |
| serkku | 67 | 0,3 | lapsi | 38 | 0,2 |

Liite 4: Koko aineiston yleisimmät verbit (n=101)

| <u>lekseemi</u> | <u>kpl</u> | <u>% verbeistä</u> | | | |
|-----------------|------------|--------------------|------------|----|-----|
| olla | 5347 | 22,5 | nousta | 61 | 0,3 |
| mennä | 1823 | 7,7 | hypätä | 59 | 0,2 |
| saada | 1006 | 4,2 | saapua | 59 | 0,2 |
| ei | 802 | 3,4 | ajatella | 56 | 0,2 |
| lähteä | 764 | 3,2 | istua | 56 | 0,2 |
| tulla | 748 | 3,1 | pukea | 56 | 0,2 |
| syödä | 572 | 2,4 | tapahtua | 55 | 0,2 |
| haluta | 467 | 2,0 | katsella | 53 | 0,2 |
| ostaa | 396 | 1,7 | laskea | 52 | 0,2 |
| käydä | 355 | 1,5 | palata | 52 | 0,2 |
| voida | 335 | 1,4 | muistaa | 50 | 0,2 |
| päästä | 318 | 1,3 | osata | 49 | 0,2 |
| sanoa | 316 | 1,3 | tuoda | 49 | 0,2 |
| nukkua | 308 | 1,3 | nukahtaa | 48 | 0,2 |
| herätä | 298 | 1,3 | tarvita | 48 | 0,2 |
| uida | 269 | 1,1 | viettää | 47 | 0,2 |
| tehdä | 263 | 1,1 | loppua | 45 | 0,2 |
| katsoa | 262 | 1,1 | pyytää | 44 | 0,2 |
| pelata | 246 | 1,0 | lasketella | 43 | 0,2 |
| alkaa | 243 | 1,0 | vaihtaa | 41 | 0,2 |
| pitää | 213 | 0,9 | muuttaa | 40 | 0,2 |
| ottaa | 189 | 0,8 | hoitaa | 39 | 0,2 |
| voittaa | 167 | 0,7 | kestää | 39 | 0,2 |
| nähdä | 162 | 0,7 | kuulla | 39 | 0,2 |
| kysyä | 119 | 0,5 | pakata | 38 | 0,2 |
| toivoa | 119 | 0,5 | kuulua | 37 | 0,2 |
| leikkiä | 118 | 0,5 | paistaa | 36 | 0,2 |
| hakea | 117 | 0,5 | pestä | 36 | 0,2 |
| antaa | 111 | 0,5 | näyttää | 35 | 0,1 |
| lukea | 108 | 0,5 | tavata | 35 | 0,1 |
| odottaa | 104 | 0,4 | tykätä | 34 | 0,1 |
| ajaa | 103 | 0,4 | kulua | 33 | 0,1 |
| viedä | 103 | 0,4 | matkustaa | 33 | 0,1 |
| laittaa | 88 | 0,4 | soida | 33 | 0,1 |
| huomata | 87 | 0,4 | toteutua | 33 | 0,1 |
| tietää | 84 | 0,4 | etsiä | 32 | 0,1 |
| soittaa | 82 | 0,3 | tahtoa | 32 | 0,1 |
| juosta | 77 | 0,3 | mieltiä | 31 | 0,1 |
| ratsastaa | 76 | 0,3 | joutua | 30 | 0,1 |
| kertoa | 75 | 0,3 | shoppailla | 30 | 0,1 |
| päättää | 72 | 0,3 | hyppiä | 29 | 0,1 |
| lentää | 71 | 0,3 | jännittää | 29 | 0,1 |
| vastata | 70 | 0,3 | kuunnella | 29 | 0,1 |
| avata | 67 | 0,3 | jatkaa | 28 | 0,1 |
| huutaa | 67 | 0,3 | laulaa | 28 | 0,1 |
| jäädä | 65 | 0,3 | ihmetellä | 27 | 0,1 |
| kävellä | 65 | 0,3 | tuntua | 27 | 0,1 |
| juoda | 63 | 0,3 | valita | 27 | 0,1 |
| löytää | 63 | 0,3 | aikoa | 26 | 0,1 |
| ruveta | 63 | 0,3 | puhua | 26 | 0,1 |
| asua | 61 | 0,3 | | | |

Liite 5: Koko aineiston yleisimmät adjektiivit (n=101)

| <u>lekseemi</u> | <u>kpl</u> | <u>% adjektiiveista</u> | | | |
|-----------------|------------|-------------------------|-------------|----|-----|
| kiva | 237 | 5,5 | semmoinen | 13 | 0,3 |
| hyvä | 207 | 4,8 | sininen | 13 | 0,3 |
| oma | 191 | 4,5 | viime | 13 | 0,3 |
| iso | 174 | 4,1 | kuuluisa | 12 | 0,3 |
| paras | 169 | 3,9 | punainen | 12 | 0,3 |
| uusi | 165 | 3,9 | tuttu | 12 | 0,3 |
| ihana | 140 | 3,3 | kallis | 11 | 0,3 |
| hauska | 133 | 3,1 | kesäinen | 11 | 0,3 |
| hieno | 126 | 2,9 | rauhallinen | 11 | 0,3 |
| seuraava | 110 | 2,6 | syvä | 11 | 0,3 |
| pieni | 105 | 2,5 | tylsä | 11 | 0,3 |
| sellainen | 96 | 2,2 | vapaa | 11 | 0,3 |
| pitkä | 91 | 2,1 | entinen | 10 | 0,2 |
| lämmin | 78 | 1,8 | kirkas | 10 | 0,2 |
| kaunis | 63 | 1,5 | vaikea | 10 | 0,2 |
| tosi | 60 | 1,4 | ikioma | 9 | 0,2 |
| suuri | 58 | 1,4 | jännittävä | 9 | 0,2 |
| mukava | 50 | 1,2 | näköinen | 9 | 0,2 |
| iloinen | 48 | 1,1 | paha | 9 | 0,2 |
| kuuma | 44 | 1,0 | varma | 9 | 0,2 |
| onnellinen | 42 | 1,0 | viileä | 9 | 0,2 |
| valmis | 40 | 0,9 | hyvä | 8 | 0,2 |
| musta | 37 | 0,9 | kultainen | 8 | 0,2 |
| kylmä | 32 | 0,7 | köyhä | 8 | 0,2 |
| vanha | 32 | 0,7 | pitkään | 8 | 0,2 |
| aurinkoinen | 28 | 0,7 | tyytyväinen | 8 | 0,2 |
| viimeinen | 28 | 0,7 | valtava | 8 | 0,2 |
| oikea | 27 | 0,6 | virkeä | 8 | 0,2 |
| valkoinen | 26 | 0,6 | väsynyt | 8 | 0,2 |
| upea | 25 | 0,6 | aika | 7 | 0,2 |
| korkea | 23 | 0,5 | ensimmäinen | 7 | 0,2 |
| outo | 23 | 0,5 | hiljainen | 7 | 0,2 |
| rikas | 23 | 0,5 | kokoinen | 7 | 0,2 |
| kaikenlainen | 22 | 0,5 | kova | 7 | 0,2 |
| tavallinen | 21 | 0,5 | lyhyt | 7 | 0,2 |
| suora | 20 | 0,5 | nuori | 7 | 0,2 |
| ilmainen | 19 | 0,4 | pehmeä | 7 | 0,2 |
| koko | 19 | 0,4 | raikas | 7 | 0,2 |
| erilainen | 17 | 0,4 | samanlainen | 7 | 0,2 |
| nopea | 17 | 0,4 | täysi | 7 | 0,2 |
| eri | 16 | 0,4 | villi | 7 | 0,2 |
| kauhea | 16 | 0,4 | herkullinen | 6 | 0,1 |
| ruskea | 16 | 0,4 | läheinen | 6 | 0,1 |
| kiltti | 15 | 0,4 | myöhä | 6 | 0,1 |
| söpö | 15 | 0,4 | puhdas | 6 | 0,1 |
| tällainen | 15 | 0,4 | sopiva | 6 | 0,1 |
| huono | 14 | 0,3 | suomalainen | 6 | 0,1 |
| hurja | 14 | 0,3 | tyhjä | 6 | 0,1 |
| mahtava | 14 | 0,3 | täydellinen | 6 | 0,1 |
| pimeä | 14 | 0,3 | vihreä | 6 | 0,1 |
| mahdollinen | 13 | 0,3 | | | |

Liite 6: Koko aineiston yleisimmät adverbit (n=101)

| <u>lekseemi</u> | <u>kpl</u> | <u>% adverbeistä</u> | | | |
|-----------------|------------|----------------------|-----------|----|-----|
| sitten | 1089 | 12,1 | huomenna | 35 | 0,4 |
| niin | 543 | 6,1 | nopeasti | 35 | 0,4 |
| siellä | 476 | 5,3 | kiinni | 34 | 0,4 |
| paljon | 306 | 3,4 | oikein | 33 | 0,4 |
| myös | 230 | 2,6 | mukaan | 31 | 0,3 |
| vielä | 226 | 2,5 | koti | 30 | 0,3 |
| jo | 210 | 2,3 | lopulta | 30 | 0,3 |
| nyt | 150 | 1,7 | vaan | 30 | 0,3 |
| heti | 149 | 1,7 | aikaisin | 29 | 0,3 |
| vain | 125 | 1,4 | sisään | 29 | 0,3 |
| tosi | 114 | 1,3 | täynnä | 29 | 0,3 |
| Silloin | 110 | 1,2 | ympäri | 28 | 0,3 |
| ulos | 105 | 1,2 | jolloin | 27 | 0,3 |
| sinne | 103 | 1,1 | siis | 26 | 0,3 |
| aina | 99 | 1,1 | sisällä | 26 | 0,3 |
| kotona | 97 | 1,1 | alas | 24 | 0,3 |
| takaisin | 97 | 1,1 | sillä | 24 | 0,3 |
| ihan | 95 | 1,1 | vihdoin | 24 | 0,3 |
| kerran | 94 | 1,0 | edes | 23 | 0,3 |
| taas | 89 | 1,0 | enemmän | 23 | 0,3 |
| tänään | 83 | 0,9 | kuitenkin | 23 | 0,3 |
| todella | 81 | 0,9 | ylös | 23 | 0,3 |
| sieltä | 76 | 0,8 | onneksi | 21 | 0,2 |
| vähän | 76 | 0,8 | yli | 21 | 0,2 |
| kyllä | 74 | 0,8 | äkkiä | 21 | 0,2 |
| pois | 74 | 0,8 | kohta | 20 | 0,2 |
| hyvin | 68 | 0,8 | koskaan | 20 | 0,2 |
| noin | 68 | 0,8 | luo | 20 | 0,2 |
| melkein | 66 | 0,7 | myöhemmin | 20 | 0,2 |
| vasta | 63 | 0,7 | ikinä | 19 | 0,2 |
| ensin | 61 | 0,7 | kuinka | 19 | 0,2 |
| päälle | 60 | 0,7 | näin | 19 | 0,2 |
| pian | 58 | 0,6 | siksi | 19 | 0,2 |
| aivan | 56 | 0,6 | vieressä | 19 | 0,2 |
| yhtäkkiä | 55 | 0,6 | luona | 18 | 0,2 |
| ehkä | 54 | 0,6 | lähellä | 18 | 0,2 |
| perillä | 53 | 0,6 | silti | 18 | 0,2 |
| juuri | 52 | 0,6 | vastaan | 18 | 0,2 |
| kauan | 51 | 0,6 | yhdessä | 18 | 0,2 |
| enää | 49 | 0,5 | edessä | 17 | 0,2 |
| joskus | 49 | 0,5 | ennen | 17 | 0,2 |
| ulkona | 48 | 0,5 | kuluttua | 17 | 0,2 |
| sisälle | 45 | 0,5 | luokse | 17 | 0,2 |
| tietenkin | 45 | 0,5 | nimittäin | 17 | 0,2 |
| ainakin | 44 | 0,5 | päin | 17 | 0,2 |
| kunnes | 43 | 0,5 | tietysti | 17 | 0,2 |
| kohti | 39 | 0,4 | viimein | 17 | 0,2 |
| varmaan | 37 | 0,4 | hieman | 16 | 0,2 |
| yksin | 37 | 0,4 | uudestaan | 16 | 0,2 |
| menossa | 36 | 0,4 | yhtään | 16 | 0,2 |
| aika | 35 | 0,4 | | | |

Liite 7: 2.-, 4.- ja 6.-luokkalaisten kirjoitelmien sata yleisintä lekseemiä (kpl).

| | 2. luokka | | 4. luokka | | 6. luokka | |
|----|-----------|-----|-----------|------|-----------|------|
| 1 | ja | 676 | olla | 1222 | olla | 1074 |
| 2 | olla | 666 | ja | 1067 | ja | 971 |
| 3 | minä | 301 | minä | 491 | se | 358 |
| 4 | mennä | 262 | mennä | 486 | mennä | 339 |
| 5 | se | 240 | se | 460 | minä | 261 |
| 6 | saada | 196 | kun | 337 | kun | 243 |
| 7 | sitten | 195 | sitten | 260 | päivä | 228 |
| 8 | kun | 164 | me | 247 | sitten | 180 |
| 9 | me | 159 | että | 214 | lähteä | 175 |
| 10 | päivä | 149 | päivä | 209 | me | 172 |
| 11 | että | 132 | tulla | 201 | että | 169 |
| 12 | haluta | 124 | saada | 198 | ei | 169 |
| 13 | ei | 117 | ei | 167 | joka | 164 |
| 14 | äiti | 98 | lähteä | 155 | saada | 146 |
| 15 | tulla | 93 | syödä | 140 | tulla | 135 |
| 16 | lähteä | 89 | niin | 133 | syödä | 133 |
| 17 | hän | 88 | äiti | 122 | kanssa | 122 |
| 18 | syödä | 88 | mutta | 121 | kaikki | 112 |
| 19 | niin | 86 | kanssa | 121 | jälkeen | 111 |
| 20 | joka | 78 | joka | 114 | mutta | 110 |
| 21 | kanssa | 74 | jälkeen | 108 | unelma | 105 |
| 22 | mutta | 72 | unelma | 98 | siellä | 99 |
| 23 | koti | 65 | siellä | 96 | hän | 99 |
| 24 | päästä | 64 | kaikki | 94 | luku | 96 |
| 25 | koira | 63 | herätä | 92 | päästä | 92 |
| 26 | siellä | 61 | nukkua | 92 | niin | 86 |
| 27 | sanoa | 53 | koti | 86 | ostaa | 86 |
| 28 | kaikki | 51 | isä | 81 | voida | 84 |
| 29 | unelma | 50 | päästä | 81 | käydä | 73 |
| 30 | koulu | 48 | sanoa | 81 | äiti | 71 |
| 31 | isä | 47 | ostaa | 79 | nukkua | 68 |
| 32 | mikä | 46 | katsoa | 77 | aamu | 66 |
| 33 | aamu | 46 | käydä | 74 | kello | 66 |
| 34 | paljon | 44 | koira | 71 | mikä | 65 |
| 35 | kaveri | 44 | aamu | 70 | alkaa | 65 |
| 36 | herätä | 43 | koska | 69 | koti | 64 |
| 37 | käydä | 43 | alkaa | 67 | herätä | 63 |
| 38 | ostaa | 42 | mikä | 66 | paljon | 61 |
| 39 | koko | 40 | hän | 62 | vielä | 60 |
| 40 | kiva | 39 | tehdä | 62 | koska | 60 |
| 41 | kissa | 38 | paljon | 59 | hotelli | 60 |
| 42 | ne | 36 | kaveri | 58 | aika | 59 |
| 43 | nimi | 36 | jo | 57 | kaveri | 58 |
| 44 | pelata | 36 | pitää | 57 | myös | 54 |
| 45 | kerran | 35 | vielä | 56 | jo | 53 |
| 46 | nukkua | 35 | haluta | 56 | tehdä | 53 |
| 47 | voida | 35 | tunti | 55 | hyvä | 52 |
| 48 | oma | 34 | oma | 52 | ilta | 52 |
| 49 | uida | 32 | pelata | 52 | uida | 52 |
| 50 | leikkiä | 31 | voida | 50 | voittaa | 52 |
| 51 | tehdä | 31 | nimi | 49 | matka | 51 |

| | 2. luokka | 4. luokka | 6. luokka |
|-----|--------------|--------------|---------------|
| 52 | toivoa | 31 uida | 49 ystävä |
| 53 | katsoa | 30 myös | 48 paras |
| 54 | koska | 28 kello | 48 jokin |
| 55 | pieni | 27 ilta | 47 tunti |
| 56 | myös | 26 hevonen | 44 katsoa |
| 57 | jälkeen | 26 heti | 42 ottaa |
| 58 | he | 26 aika | 42 haluta |
| 59 | hevonen | 26 koulu | 42 hieno |
| 60 | loppu | 26 ottaa | 42 iso |
| 61 | iso | 25 iso | 41 sanoa |
| 62 | raha | 25 yksi | 41 nyt |
| 63 | nähdä | 25 aamupala | 41 muu |
| 64 | hauska | 24 loppu | 41 isä |
| 65 | uusi | 24 kuin | 40 kiva |
| 66 | aina | 24 huone | 40 ne |
| 67 | Liisa | 24 palvelija | 40 raha |
| 68 | tosi | 23 kiva | 39 tämä |
| 69 | eräs | 22 tosi | 39 tai |
| 70 | hyvä | 21 auto | 39 he |
| 71 | vielä | 21 hyvä | 38 auto |
| 72 | ruoka | 21 ihana | 38 kellonaika |
| 73 | aamupala | 20 antaa | 38 pelata |
| 74 | karkki | 20 nähdä | 37 pitää |
| 75 | tyttö | 20 he | 36 jos |
| 76 | voittaa | 20 ulos | 35 pitkä |
| 77 | jos | 19 uusi | 34 vain |
| 78 | kauppa | 19 kaksi | 34 kuin |
| 79 | alkaa | 19 ne | 34 nähdä |
| 80 | ottaa | 19 karkki | 34 oma |
| 81 | paras | 18 odottaa | 34 vaikka |
| 82 | pitää | 18 päälle | 33 yksi |
| 83 | silloin | 17 jos | 33 itse |
| 84 | ulos | 17 jokin | 32 aurinko |
| 85 | vaikka | 17 sisko | 32 koulu |
| 86 | kaksi | 17 hauska | 31 tosi |
| 87 | yksi | 17 pieni | 30 heti |
| 88 | koiranpentu | 17 taas | 30 ranta |
| 89 | lahja | 17 tämä | 30 aamupala |
| 90 | syntymäpäivä | 17 laittaa | 30 koko |
| 91 | talo | 17 voittaa | 30 seuraava |
| 92 | asua | 17 koko | 29 todella |
| 93 | kysyä | 17 hakea | 29 Helsinki |
| 94 | jo | 16 lukea | 29 odottaa |
| 95 | kotona | 16 vaikka | 28 palata |
| 96 | nyt | 16 moni | 28 noin |
| 97 | vähän | 16 matka | 28 takaisin |
| 98 | aika | 16 sänky | 28 koira |
| 99 | isi | 16 ystävä | 28 perhe |
| 100 | maailma | 16 hieno | 27 vaate |
| 101 | osata | 16 seuraava | 27 soittaa |
| 102 | | nyt | 27 viedä |
| 103 | | silloin | 27 |
| 104 | | vain | 27 |
| 105 | | hotelli | 27 |
| 106 | | ruoka | 27 |